

Phase-space-simulation method for quantum computation with magic states on qubitsRobert Raussendorf^{1,2}, Juani Bermejo-Vega^{3,*}, Emily Tyhurst⁴, Cihan Okay^{1,2} and Michael Zurek^{1,2}¹*Department of Physics & Astronomy, University of British Columbia, Vancouver, British Columbia, Canada V6T1Z1*²*Stewart Blusson Quantum Matter Institute, University of British Columbia, Vancouver, British Columbia, Canada V6T1Z4*³*Dahlem Center for Complex Quantum Systems, Physics Department, Freie Universität Berlin, 14195 Berlin, Germany*⁴*Department of Physics, University of Toronto, Toronto, Ontario, Canada M5S 1A7*

(Received 21 June 2019; published 31 January 2020)

We propose a method for classical simulation of finite-dimensional quantum systems, based on sampling from a quasiprobability distribution, i.e., a generalized Wigner function. Our construction applies to all finite dimensions, with the most interesting case being that of qubits. For multiple qubits, we find that quantum computation by Clifford gates and Pauli measurements on magic states can be efficiently classically simulated if the quasiprobability distribution of the magic states is non-negative. This provides the so far missing qubit counterpart of the corresponding result [V. Veitch *et al.*, *New J. Phys.* **14**, 113011 (2012)] applying only to odd dimension. Our approach is more general than previous ones based on mixtures of stabilizer states. Namely, all mixtures of stabilizer states can be efficiently simulated, but for any number of qubits there also exist efficiently simulable states outside the stabilizer polytope. Further, our simulation method extends to negative quasiprobability distributions, where it provides probability estimation. The simulation cost is then proportional to a robustness measure squared. For all quantum states, this robustness is smaller than or equal to robustness of magic.

DOI: [10.1103/PhysRevA.101.012350](https://doi.org/10.1103/PhysRevA.101.012350)**I. INTRODUCTION**

How to mark the classical-to-quantum boundary is a question that dates back almost to the beginning of quantum theory. Ehrenfest's theorem [1] provides an early insight, and the Einstein-Podolsky-Rosen paradox [2] and Schrödinger's cat [3] are two early puzzles. The advent of quantum computation [4–6] added a computational angle: When does it become hard to simulate a quantum mechanical computing device on a classical computer? Which quantum mechanical resource do quantum computers harness to generate a computational speedup?

One instructive computational model is quantum computation with magic states (QCM) [7]. In QCM, both “traditional” indicators of quantumness (developed in the fields of quantum optics and foundations of quantum mechanics) and a computational indicator can be applied. From quantum optics and foundations, the indicators are the negativity of a Wigner function [8–10] and the breakdown of noncontextual hidden variable models [11–13]. Computer science is concerned with the breakdown of efficient classical simulation.

In the particular setting of QCM, an important distinction arises between the cases of even and odd local Hilbert space dimension d . If d is odd, then all three of the above indicators for the classical-to-quantum boundary align [14–16]. This is a very satisfying situation: the physicist, the philosopher, and the computer scientist can have compatible notions of what is “quantum.”

In even local dimension, the situation differs starkly. Noncontextual hidden variable models for QCM are not viable regardless of computational power [16], which voids the

foundational indicator, and furthermore obstructs the view of contextuality as a computational resource. Also, the multi-qubit Wigner functions constructed to date do not support efficient classical simulation of QCM by sampling over phase space. Thus, the physics and computer-science-based criteria for classicality differ, which is an unsatisfactory state of affairs compared to odd d . The purpose of this paper is to align the perspectives of the physicist and the computer scientist on the classical-to-quantum transition in QCM on qubits.

To prepare for the subsequent discussion, we provide a short summary of QCM and the role of the Wigner function in it. Quantum computation with magic states operates with a restricted set of instructions, the Clifford gates. These are unitary operations defined by the property that they map all Pauli operators onto Pauli operators under conjugation. Clifford gates are not universal and, in fact, can be efficiently classically simulated [17]. This operational restriction is compensated for by invoking the “magic” states, which are special quantum states that cannot be created by Clifford gates and Pauli measurements. Suitable magic states restore quantum computational universality; and in fact QCM is a leading paradigm for fault-tolerant universal quantum computation. In sum, computational power is transferred from the quantum gates to the magic states, and one is thus led to ask: Which quantum properties give the magic states their computational power?

One such property is, for odd d at least, negativity in the Wigner function. A quantum speedup can arise only if the Wigner function of the magic states assumes negative values. If, to the contrary, the Wigner function is positive, then the whole quantum computation can be efficiently classically simulated [14,15]. Further, a positive Wigner function is, for $n \geq 2$ quantum systems, equivalent to the existence of a noncontextual hidden variable model [16,18]. Both Wigner

*Present address: University of Granada, Av. Fuentenueva s/n., 18071 Granada, Spain.

function negativity and contextuality of the magic states are therefore necessary quantum computational resources.

As we noted, this picture applies only if the local Hilbert space dimension is odd. This excludes the full multiqubit case, which arguably is the most important. Approaches to the qubit case have been made, e.g., through the rebit scenario [19] and multiqubit settings with operational restrictions [20,21], or by invoking a Wigner function over Grassmann variables [22], or multiple Wigner functions at once [23]. Common to these approaches is that, unlike for odd d [14], they do not efficiently simulate the evolution under general Clifford gates and Pauli measurements by sampling, a.k.a. weak simulation [24–26].

An alternative approach to weak simulation is by defining a quasiprobability function over stabilizer states [7,27,28], bypassing Wigner functions. It has the advantage of efficiently simulating all Clifford circuits on positively represented states. For multiqubit systems, it has so far been unknown how the stabilizer method relates to Wigner functions, but we clarify the relation here.

In this paper, we provide the thus far missing phase space picture for QCM on multiqubit systems. Central to our discussion is a quasiprobability function defined for all local Hilbert space dimensions d and all numbers of subsystems n . When applied to odd d , it reproduces the known finite-dimensional adaptation [29–31] of the original Wigner function [8]; but for even d , in particular $d = 2$, it is different. Then this quasiprobability function requires a phase space of increased size, in accordance with [32]. Even in $d = 2$, the positivity of this quasiprobability is preserved under all Pauli measurements. This property is crucial for the efficient classical simulation of QCM on positively represented states. Also, this simulation contains the efficient classical simulation [7] of stabilizer mixtures as a special case. We thus reproduce the essential features of the odd-dimensional scenario in $d = 2$.

Starting from the definition of the quasiprobability function W , we treat the following subjects: characterization of phase space for $d = 2$, preservation of positivity of W under Pauli measurements, covariance of W under all Clifford unitaries, efficient classical simulation of QCM for $W \geq 0$, relation to the qubit stabilizer formalism, hardness of classical simulation for $W < 0$, and a monotone under the free operations.

In summary, we arrive at a description that resembles the corresponding scenario in odd local dimension. Namely, negativity in the quasiprobability distribution W for the initial magic state is a necessary precondition for quantum speedup. However, one difference between even and odd d remains. In odd d , every positive Wigner function is also a noncontextual hidden variable model. This is not so for even d , due to the phenomenon of state-independent contextuality among Pauli observables.

II. RESULTS AND OUTLINE

A. Summary of results

This paper addresses the full n -qubit case of quantum computation with magic states, from the perspectives of the classical-to-quantum transition and quantum computational

resources. For the case of local dimension $d = 2$ we closely reproduce the relations between Wigner function and efficient classical simulation existing in odd d . Central to our discussion is a quasiprobability function W defined for all local Hilbert space dimensions d . It has the following general properties:

- (i) For all n and d , W is Clifford-covariant and positivity-preserving under Pauli measurements.
- (ii) If the local Hilbert space dimension d is even, W_ρ is nonunique for any given quantum state ρ . The set of phase point operators corresponding to W is overcomplete.
- (iii) If d is odd and $n \geq 2$, then W reduces to the standard Wigner function [29,30] for odd finite dimension.
- (iv) For all n and d , the stabilizer formalism is contained as a special case. All stabilizer states can be positively represented by W and efficiently updated under Clifford operations.
- (v) The present description goes beyond the stabilizer formalism. In particular, for $d = 2$, for every number n of qubits there exist nonmixtures of stabilizer states which are positively represented by W . Furthermore, for any quantum state ρ , the one-norm of the optimal W_ρ is smaller than or equal to the robustness of magic $\mathfrak{R}_S(\rho)$. (Both robustness measures are instances of sum negativity [28].)

The following properties of W for special values of n (and $d = 2$) are also worth noting. (a) The eight-state model [33] is a special case of W , namely, for $n = 1$. (b) For Mermin’s square [13], the present simulation algorithm saturates the lower bound [34] on the memory cost of classical simulation. (c) Up to two copies of magic T and H states are positively represented by W .

We establish the following main results: (I) The set of states positively represented by W is closed under Pauli measurement (Theorem 2 in Sec. V). (II) If a quantum state ρ has a non-negative function W_ρ , and W_ρ can be efficiently sampled, then, for every Clifford circuit applied to ρ , the corresponding measurement statistics can be efficiently sampled (Theorem 3 in Sec. VI). In this sense, $W \geq 0$ leads to efficient classical simulation of the corresponding quantum computation. (III) For $d = 2$, $n \geq 2$, the n -system phase space has a more complicated structure than in the case of odd d , reflecting the fact that the phase point operators are dependent. The points in generalized multiqubit phase space are classified (Theorem 1 in Sec. IV). (IV) There exists a robustness measure \mathfrak{R} which bounds the hardness of classical simulation of quantum computation with magic states, when $W_{\rho_{\text{init}}} < 0$ for the initial state ρ_{init} . \mathfrak{R} is less than or equal to the robustness of magic (Lemma 9), and a monotone under Clifford unitaries and Pauli measurements (Theorem 4 in Sec. VII).

B. Outline

The remainder of this paper is organized as follows. In Sec. III we define a quasiprobability function W . We show that it reduces to Gross’ Wigner function [29] whenever the local Hilbert space dimension d is odd but, more importantly, is different in even dimension. Specifically, W represents all quantum states redundantly for even d , which enables Clifford covariance and positivity preservation under Pauli measurement. In Sec. IV we analyze the structure of the phase space on which W lives for the case of multiple qubits. In particular,

we classify the points of phase space. We also clarify the relation to the qubit stabilizer states and their mixtures.

In Secs. V and VI we turn to dynamics. In Sec. V we discuss the update of W under Pauli measurement, and in Sec. VI the efficient classical simulation of QCM for positive W .

In Sec. VII we address the case of $W_\rho < 0$. We discuss hardness of classical simulation, as well as the elements of a resource theory based on W .

In Sec. VIII we discuss the extent to which the quasiprobability function W satisfies the Stratonovich-Weyl criteria, and its relation to hidden variable models. We conclude in Sec. IX.

III. THE QUASIPROBABILITY FUNCTION

In this section we introduce the generalized n -qudit phase space \mathcal{V} , for any local Hilbert space dimension d , and a quasiprobability distribution $W : \mathcal{V} \rightarrow \mathbb{R}$ living on it. In Sec. III A we define the phase point operators corresponding to W , and in Sec. III B identify a minimal set of them. Section III C reveals the cohomological underpinning of our construction, which links the present subject to parity proofs of quantum contextuality [35] and contextuality in measurement-based quantum computation [36].

A. Generalized phase space

We choose a phase convention for the Pauli operators,

$$T_a = e^{i\phi(a)} X(a_X) Z(a_Z), \quad \forall a = (a_X, a_Z) \in E := \mathbb{Z}_d^{2n}. \quad (1)$$

Therein, the function $\phi : E \rightarrow \mathbb{R}$ has to satisfy the constraint that $(T_a)^d = I$, for all $a \in E$. As a consequence of this condition, all eigenvalues of the operators T_a are of the form ω^k , $k \in \mathbb{N}$, with $\omega := \exp(2\pi i/d)$.

We now proceed to the definition of the phase point operators. We consider a subset Ω of E and a function $\gamma : \Omega \rightarrow \mathbb{Z}_d$, both subject to additional constraints that will be specified in Definitions 2–4 below. The pair (Ω, γ) specifies a corresponding phase point operator A_Ω^γ ,

$$A_\Omega^\gamma := \frac{1}{d^n} \sum_{b \in \Omega} \omega^{\gamma(b)} T_b, \quad (2)$$

with the constraint that

$$\omega^{\gamma(0)} T_0 = I. \quad (3)$$

When comparing Eq. (2) to the phase point operators of the previously discussed qudit [14], rebit [19], and restricted qubit [20] cases, we note that the overall structure remains the same. In this case, the sets Ω are an additional varying parameter, and the phase space thereby becomes larger.

Based on the phase point operators A_Ω^γ of Eq. (2), we introduce the counterpart to the Wigner function that applies to our setting. The generalized phase space \mathcal{V} consists of all admissible pairs (Ω, γ) , to be specified below. Any n -system quantum state ρ can be expanded in terms of a function $W_\rho : \mathcal{V} \rightarrow \mathbb{R}$,

$$\rho = \sum_{(\Omega, \gamma) \in \mathcal{V}} W_\rho(\Omega, \gamma) A_\Omega^\gamma. \quad (4)$$

The reason for imposing Eq. (3) is that it implies $\text{Tr} A_\Omega^\gamma = 1$, for all $(\Omega, \gamma) \in \mathcal{V}$. Hence, W defined in Eq. (4) is a quasiprobability distribution. As we see shortly, it generalizes the Wigner function [29] for odd-dimensional qudits to qubits.

We note that when d is even, the quasiprobability distribution W_ρ is nonunique because the set of phase point operators of Eq. (2) is overcomplete.

Definition 1. An n -qudit quantum state ρ is *positively representable* if it can be expanded in the form of Eq. (4), with $W_\rho(\Omega, \gamma) \geq 0$, for all $(\Omega, \gamma) \in \mathcal{V}$.

The efficient classical simulation algorithm described in Sec. VI applies to positively representable quantum states ρ . The nonuniqueness of W_ρ allows for more positively representable states than prior quasiprobability representations.

We now turn to the properties of admissible sets Ω and functions γ that define points in the phase space \mathcal{V} . To begin, we define a function β which encodes how translation operators on phase space compose,

$$T_a T_b = \omega^{\beta(a,b)} T_{a+b}, \quad \forall a, b \in E : [T_a, T_b] = 0. \quad (5)$$

We further define the symplectic product

$$[a, b] := a_X b_Z - a_Z b_X \quad \text{mod } d, \quad (6)$$

and hence $[a, b] = 0 \iff [T_a, T_b] = 0$.

The function β satisfies the relation

$$\beta(a, b) + \beta(a + b, c) - \beta(b, c) - \beta(a, b + c) = 0 \quad \text{mod } d, \quad (7)$$

for $a, b, c \in E$. We state this relation for later reference. It is a consequence of the associativity of operator multiplication. Consider the operator product $T_a T_b T_c = T_a (T_b T_c) = (T_a T_b) T_c$, and expand $T_a (T_b T_c) = \omega^{\beta(a,b+c)+\beta(b,c)} T_{a+b+c}$, $(T_a T_b) T_c = \omega^{\beta(a,b)+\beta(a+b,c)} T_{a+b+c}$. Comparing the two equivalent expressions yields Eq. (7).

Then, it follows straightforwardly from the definition Eq. (5) of β that

$$\beta(a, b) = \beta(b, a), \quad \forall a, b \text{ with } [a, b] = 0. \quad (8)$$

We constrain Ω by the following definitions:

Definition 2. A set $\Omega \subset E$ is *closed under inference* if it holds that

$$a, b \in \Omega \wedge [a, b] = 0 \implies a + b \in \Omega. \quad (9)$$

The motivation for this definition is that if T_a and T_b can be simultaneously measured, then the value of T_{a+b} can be inferred from the measurement outcomes, through relation (5). A consequence of the closedness under inference is that $0 \in \Omega$ for all closed sets Ω .

Definition 3. A set $\Omega \subset E$ is *noncontextual* if there exists a value assignment $\gamma : \Omega \rightarrow \mathbb{Z}_d$ that satisfies the condition

$$\gamma(a) + \gamma(b) - \gamma(a + b) = \beta(a, b), \quad (10)$$

for all $a, b \in \Omega$, and $[a, b] = 0$.

To motivate the nomenclature, if the set $\Omega \subset E$ is noncontextual per the above definition, then it does not admit a parity-based contextuality proof [35]. Namely, Eq. (10) represents the constraints on noncontextual value assignments γ that result from the operator constraints Eq. (5). If these constraints can be satisfied, then there is no parity-based contextuality proof.

Definition 4. The generalized phase space \mathcal{V} consists of all pairs (Ω, γ) such that (i) Ω is closed under inference, (ii) Ω is noncontextual, (iii) $\gamma : \Omega \rightarrow \mathbb{Z}_d$ satisfies the relation Eq. (10), and (iv) Eq. (3) holds.

Thus, for the generalized phase space \mathcal{V} , the only sets Ω that matter are simultaneously closed and noncontextual. For short, we call such sets ‘‘cnc.’’

B. Maximal sets Ω

The cnc sets Ω partially specify the points in phase space, and it is thus desirable to eliminate possible redundancies among them. It turns out that only the ‘‘maximal’’ sets Ω need to be considered for \mathcal{V} .

Definition 5. A cnc set $\Omega \subset E$ is maximal if there is no cnc set $\tilde{\Omega} \subset E$ such that $\Omega \subsetneq \tilde{\Omega}$.

We denote by \mathcal{V}_M the subset of \mathcal{V} constructed only from the maximal cnc sets Ω . Then any quantum state ρ has expansions like Eq. (4), but with \mathcal{V} replaced by \mathcal{V}_M . If one of those expansions is non-negative, then we say that ρ is positively representable w.r.t. \mathcal{V}_M .

Lemma 1. For any n and d , a quantum state ρ is positively representable w.r.t. \mathcal{V} if and only if it is positively representable w.r.t. \mathcal{V}_M .

From the perspective of positive representability, we may therefore shrink \mathcal{V} to \mathcal{V}_M without loss. We make use of this property when discussing the case of odd d in Sec. IV A right below, and in the classification of cnc sets Ω for the multiqubit case in Sec. IV C. The proof of Lemma 1 is given in Appendix A.

C. The cohomological viewpoint

The above Definitions 3 and 4 have a cohomological underpinning, which connects the subject of the present paper to the topological treatment of parity-based contextuality proofs [35], and of contextuality in measurement-based quantum computation [36].

The cohomological picture arises as follows. The partial value assignments γ and the function β are cochains in a chain complex, with Eqs. (7) and (10) constraining them. Equation (7) says that β is a special cochain, namely, a cocycle. Now, the basic reason for why the case of even d is so much more involved than the case of odd d is that, for even d , the cocycle β is nontrivial, whereas for odd d it is trivial [35].

Equations (7) and (10) are frequently used in this paper, for example, in the update rules of the phase point operators under Pauli measurements (proof of Lemma 5), the closedness of the generalized phase space \mathcal{V} under update by Pauli measurement (proof of Lemma 7), and covariance of the quasiprobability function W under Clifford unitaries (proof of Lemma 10). These are central properties for the phase-space description of quantum computation with magic states, and they are all matters of cohomology.

The cohomological formulation is based on a chain complex \mathcal{C}_n constructed from the n -qubit Pauli operators T_a . The operator labels a define the edges of this complex; the faces of \mathcal{C}_n correspond to commuting pairs (a, b) and volumes (a, b, c) to commuting triples. For details, the interested reader is

referred to Ref. [35]. Here we only state two basic topological properties of the present scenario.

As already noted, the cochain β defined in Eq. (5) is in fact a two-cocycle, with the cocycle condition $d\beta = 0$ enforced by Eq. (7). For any given volume $v = (a, b, c)$, the coboundary $d\beta$ evaluates on v to

$$d\beta(a, b, c) := \beta(a, b) + \beta(a + b, c) - \beta(b, c) - \beta(a, b + c). \quad (11)$$

Thus, Eq. (7) says that $d\beta(v) = 0$, for all volumes v .

Equation (10) in Definition 3 also has a cohomological interpretation, namely, $d\gamma = \beta|_{\Omega \times \Omega}$, with

$$d\gamma(a, b) := \gamma(a) + \gamma(b) - \gamma(a + b), \quad (12)$$

for any face (a, b) spanned by commuting edges $a, b \in E$.

Subsequently, we use evaluations of $d\beta$ and $d\gamma$, defined in Eqs. (11) and (12), as a short-hand to express Eqs. (7) and (10). As outlined above, it is conceptually helpful to remember that $d\beta$ and $d\gamma$ denote coboundaries, but it is not required for the technical results presented in this paper.

IV. PROPERTIES OF THE PHASE SPACE \mathcal{V}

In this section, we look at the structure of the phase space \mathcal{V} more closely and make connections to previous phase space formulations. Namely, in Sec. IV A we address the relationship of this phase space to the usual qudit phase space, and in Sec. IV B we make clear the connections to the previously addressed rebit case. Further, in Sec. IV C, we classify the cnc sets Ω , and for every Ω describe the sets $\Gamma(\Omega)$ of value assignments γ . In Sec. IV D we clarify the relation to the stabilizer formalism.

A. Qudits of odd dimension

This is the only place in the present paper where we consider the case of odd d . The purpose of this section is to show that if d is odd, then for $n \geq 2$ qudits the generalized phase space \mathcal{V} reduces to the standard phase space $V = \mathbb{Z}_d^{2n}$. There the quasiprobability function W becomes the standard Wigner function [29] for odd finite-dimensional systems. Hence, the present quasiprobability function W is a generalization of the finite odd dimensional Wigner function [29], which in turn is a descendant of the original Wigner function [8].

If d is odd, then the whole set E is cnc. First, E is closed under inference by definition. And, second, it is known that in odd dimension Pauli observables have noncontextual deterministic value assignments [37,20]. These yield the functions γ , satisfying the condition Eq. (10). E is thus noncontextual.

E is furthermore the single maximal set, and, with Lemma 1, the only cnc set that needs to be considered for the phase space. Hence, the phase point operators are

$$A_E^\gamma = \frac{1}{d^n} \sum_{a \in E} \omega^{\gamma(a)} T_a,$$

with the functions γ satisfying Eqs. (3) and (10). The former condition ensures that the identity operator appears with weight $1/d^n$ in the expansion (real and positive). If $n \geq 2$, the latter condition has d^{2n} solutions for the functions γ if d is odd [18]. For a suitable choice of ϕ in Eq. (1), it holds that

$\beta \equiv 0$ (odd d only). The solutions for γ then form a vector space

$$V = \mathbb{Z}_d^{2n} \quad (\text{for odd } d).$$

We note that the case of a single qudit, $n = 1$, is an exception to the above behavior. In this case, the set \mathcal{V} has greater cardinality than \mathbb{Z}_d^2 [18].

B. Qubits and rebits

The remainder of this paper is about local Hilbert space dimension $d = 2$. This means mostly qubits, but we will occasionally also consider systems of rebits. The reason is that the major complication of the $d = 2$ case stems from Mermin's square and star [13]—two strikingly simple contextuality proofs. Those settings embed most efficiently in rebits rather than qubits, which warrants the inclusion of rebits here.

We remark that the present discussion of rebits is almost identical to the discussion of qubits, but very different from the earlier discussion of rebits in [19]. In the latter, the physically measurable observables were restricted from real Pauli operators to CSS-type Pauli operators, and the real Clifford unitaries to CSS-ness preserving Clifford unitaries. No such restrictions are imposed here. If the restriction to CSS-ness preserving operations is imposed, then Mermin's square and star, along with all other state-independent contextuality proofs based on Pauli observables, are effectively excised [19]. Here we retain those contextuality proofs, and consequently have to adjust to their presence. Notably, these contextuality proofs constrain quasiprobability distributions that preserve positivity under Pauli measurement.

We start the exploration of the $d = 2$ case with two examples that illustrate the concept of generalized phase space \mathcal{V} . The second example also illustrates the differences between contextuality, negative quasiprobability, and quantum computational power for two-level systems.

Example 1: Eight-state model. It is known that every one-qubit quantum state can be positively represented by the so-called eight-state model [33], which consists of two standard one-qubit Wigner functions tagged together. The eight-state model is an instance of the state expansion Eq. (4), namely, for $d = 2$, $n = 1$, and it contains only one set Ω ,

$$\Omega_0 = \{0, x, y, z\},$$

with $T_0 = I$, $T_x = X$, $T_y = Y$ and $T_z = Z$. It is easily checked that Ω_0 is noncontextual and closed under inference (no inference possible). The value assignments γ are constrained by Eq. (3), hence $\gamma(0) = 0$, and no constraints arise from Eq. (7) due to the lack of nontrivial commuting elements in Ω_0 . Thus, $\gamma(x)$, $\gamma(y)$ and $\gamma(z)$ can be freely chosen. There are eight resulting functions, and they define the eight states of the model.

All one-qubit quantum states can be positively represented by this model, which is strictly more than all mixtures of one-qubit stabilizer states.

Example 2: Mermin's square. Mermin's square is at the very root of the complications that arise for Wigner functions in even dimension. In particular, no n -qubit Wigner function

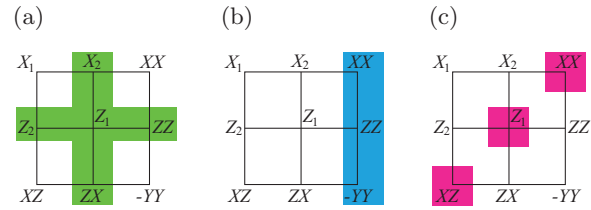


FIG. 1. Three types of cnc sets Ω for Mermin's square. (a) Union of two isotropic subspaces intersecting in one element, (b) isotropic subspace, (c) triple of anticommuting elements.

for which the corresponding phase point operators form an operator basis can preserve positivity under all Pauli measurements [20].

All observables appearing in Mermin's star are real and can thus be embedded in two rebits. Our formalism is easily adaptable to this slightly simpler scenario. Figure 1 shows three distinct types of cnc sets Ω . Type (a) is the union of two nontrivially intersecting isotropic subspaces (9 sets), type (b) is isotropic subspaces (6 sets), and type (c) is triples of anticommuting elements, i.e., one from each row and column of the square (6 sets). For each cnc set Ω of type (a), (b), and (c) of Fig. 1, the constraint Eq. (10) allows for 2^3 , 2^2 , and 2^3 functions γ , respectively. The number of phase space points of each type therefore is 72, 24, 48.

We make the following numerical observations about the two-rebit case: (i) Random sampling suggests that all 2-rebit states are positively representable; see Table III. (ii) In Fig. 2 the region of positively representable density matrices of the form

$$\rho(x, y) = \frac{1}{4}I_{12} + x(X_1X_2 + Z_1Z_2 - Y_1Y_2) + y(Z_1 + Z_2), \quad (13)$$

for $x, y \in \mathbb{R}$, is displayed for three different methods: the stabilizer method [27], the hyperoctahedral method [38], and the present phase space method. We find that all quantum states in the plane spanned by the parameters x, y are positively represented by the present phase space method, and this is not the case for the stabilizer and hyperoctahedral methods.

Example 3: Two qubits. Numerical analysis shows that two copies of the state

$$|H(\phi)\rangle := (|0\rangle + e^{-i\phi}|1\rangle)/\sqrt{2} \quad (14)$$

can be positively represented, for all angles ϕ .

C. Classification of multiqubit phase space points

Denote by $\Gamma(\Omega)$ the set of functions $\gamma : \Omega \rightarrow \mathbb{Z}_2$ that satisfy the constraints Eqs. (3) and (10). Then the following statement holds.

Lemma 2. For all sets Ω of Def. 4, $\Gamma(\Omega)$ is the coset of a vector space $U(\Omega)$.

Proof of Lemma 2. Write $\gamma = \gamma_0 + \eta$, where $\gamma_0 \in \Gamma(\Omega)$ is some reference function. Then the only condition on the functions $\eta \in U(\Omega)$ is $d\eta = 0$. Thus, if $\eta, \eta' \in U(\Omega)$ then $c\eta + c'\eta' \in U(\Omega)$, for all $c, c' \in \mathbb{Z}_2$. ■

Lemma 2 reproduces a familiar feature. In infinite and finite odd dimension, the whole phase space is an orbit under the vector space of translations. There is an origin 0 of phase space, and all other phase space points are obtained from it by

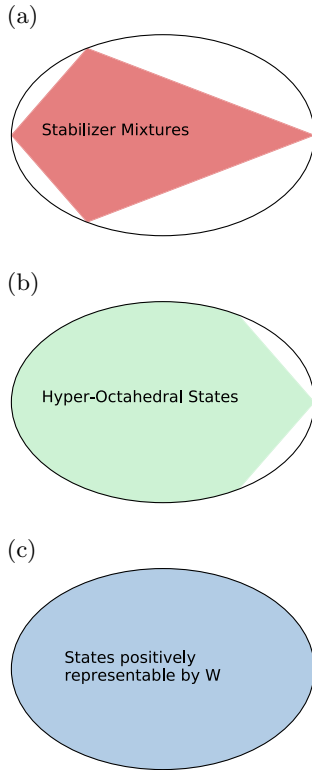


FIG. 2. Two-dimensional cross section of the two-qubit state space, as parameterized in Eq. (13). The shaded regions indicate the positively representable states by various methods; (a) mixtures of stabilizer states, (b) hyperoctahedral states [38], and (c) states positively represented by the present phase space method.

translation. In our present case of $d = 2$, the phase space \mathcal{V} splinters into many fragments, each of which corresponds to a vector space U attached to a cnc set Ω .

At this point, one question about the structure of \mathcal{V} remains: Can the cnc sets Ω be classified? It is resolved by Lemma 3 and Theorem 1 below.

Lemma 3. For n qubits, consider an isotropic subspace $\tilde{I} \subset E$ of dimension $n - m$, with $m \leq n$, and $\xi \leq 2m + 1$ elements $a_k \in E$ that pairwise anticommute but all commute with \tilde{I} . Denote $I_k := \langle a_k, \tilde{I} \rangle$ for $k = 1, \dots, \xi$. For any number n of qubits, the sets

$$\Omega = \bigcup_{k=1}^{\xi} I_k \quad (15)$$

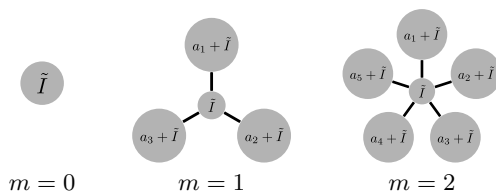


FIG. 3. Commutativity graph representation for the cosets of Eq. (15) sets. Elements pairwise commute within each vertex and elements in adjacent vertices pairwise commute. Elements in nonadjacent vertices anticommute.

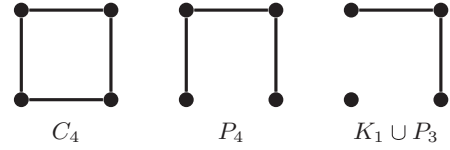


FIG. 4. Forbidden induced subgraphs of the commutativity graph, resulting from Mermin's square (also see Ref. [39]).

are noncontextual and closed under inference. Figure 3 shows a visual representation of the commutativity graph for these sets.

Proof of Lemma 3. Existence. The sets Ω of Eq. (15) exist for all m, n . To see this, consider the m -qubit Jordan-Wigner transforms of the Majorana fermion operators acting on qubits 1 to m ,

$$C_{2j-1} = I_{1..j-1} X_j Z_{j+1} Z_{j+2} \cdots Z_{m-1} Z_m, \quad (16)$$

$$C_{2j} = I_{1..j-1} Y_j Z_{j+1} Z_{j+2} \cdots Z_{m-1} Z_m,$$

for $j = 1, \dots, m$, and, if $m > 0$, the further observable

$$C_{2m+1} = Z_1 Z_2 \cdots Z_{m-1} Z_m. \quad (17)$$

Further, be \tilde{I} the isotropic subspace corresponding to a stabilizer state supported on the $n - m$ qubits numbered $m + 1, \dots, n$. Define a_k via $C_k = T_{a_k}$ as in Eqs. (16) and (17), for all $k = 1, \dots, 2m + 1$. These a_k and $a \in \tilde{I}$ have the commutation relations required.

Closedness. Consider a pair $c, d \in \Omega$ such that $[c, d] = 0$. There are two cases. (i) $c, d \in I_k$, for some k . Then $c + d \in I_k$, hence $c + d \in \Omega$.

(ii) $c \in I_k$ and $d \in I_l$, $k \neq l$. We may write $c = vx + g, d = \mu y + g'$, for some $v, \mu \in \mathbb{Z}_2$ and $g, g' \in \tilde{I}$. The commutation relation $[c, d] = 0$ then implies that $v\mu = 0$, hence either $v = 0$ or $\mu = 0$. Wlog. assume that $v = 0$. Then $c \in \tilde{I}$, hence $c, d \in I_l$. Thus, $c + d \in I_l \subset \Omega$.

In both cases, $c, d \in \Omega$ and $[c, d] = 0$ implies that $c + d \in \Omega$. Hence Ω is closed under inference.

Noncontextuality. There exists a function $\gamma|_{\tilde{I}} : \tilde{I} \rightarrow \mathbb{Z}_2$ that satisfies Eq. (10) on \tilde{I} . We now extend this function to Ω as follows. The values $\gamma(a_k)$, for $k = 1, \dots, \xi$ can be freely chosen, and for all $a \in \tilde{I}$ and all k , $\gamma(a_k + a) := \gamma(a_k) + \gamma(a) + \beta(a_k, a)$. This fully defines $\gamma : \Omega \rightarrow \mathbb{Z}_2$. All commuting triples $c, d, c + d$ lie within one of the isotropic spaces I_k forming Ω , and $d\gamma(a, b) = \beta(a, b)$ thus holds.

This establishes that the sets Ω of Eq. (15) exist for the maximum value of ξ , $\xi = 2m + 1$. One may always choose ξ smaller, which neither affects closedness nor noncontextuality. ■

Theorem 1. All maximal cnc sets Ω are of the form Eq. (15), with $\xi = 2m + 1$ and $1 \leq m \leq n$.

Proof of Theorem 1. Let $\Omega \subset E$ be closed under inference and noncontextual. We can partition the elements of Ω into two subsets, $\Omega = \{q_1, \dots, q_\mu | g_1, \dots, g_\nu\}$, where $\tilde{I} = \{g_1, \dots, g_\nu\}$ are the elements of Ω which commute with the whole set. \tilde{I} is an isotropic subspace since if two elements, a and b , commute with Ω , then clearly their sum, $a + b$, also commutes with Ω , and \tilde{I} is isotropic by definition.

If all elements of Ω pairwise commute then $\Omega = \tilde{I}$ is an isotropic subspace. Isotropic subspaces are not maximal cnc

sets because they are always contained in Eq. (15) sets with parameter $m = 1$. If Ω is not an isotropic, subspace then it can be written compactly as

$$\Omega = \bigcup_{k=1}^{\xi} \langle p_k, \tilde{I} \rangle, \tag{18}$$

where $\xi \geq 2$, the cosets $p_1 + \tilde{I}, \dots, p_{\xi} + \tilde{I}$ are distinct and q_1, \dots, q_{μ} are in the cosets $p_1 + \tilde{I}, \dots, p_{\xi} + \tilde{I}$. Note that in this form, there can be no element p_j which commutes with all of p_1, \dots, p_{ξ} because \tilde{I} is defined to contain all such elements. Now we consider the possible commutation relations that p_1, \dots, p_{ξ} can have if Ω is noncontextual.

The Mermin square is generated by products of commuting pairs of the two-qubit Pauli operators $\{X_1, X_2, Z_1, Z_2\}$. This is a contextual set. Therefore, any set which is closed under inference and contains four elements p_1, p_2, p_3, p_4 with the commutation relations like those of $\{X_1, X_2, Z_1, Z_2\}$:

$$\begin{aligned} [p_1, p_2] &= [p_1, p_4] = [p_2, p_3] = [p_3, p_4] = 0, \\ [p_1, p_3] &= [p_2, p_4] = 1 \end{aligned} \tag{19}$$

will necessarily contain the full Mermin square and therefore be contextual.

Another sufficient condition for a closed under inference set to be contextual is that it contains four elements with the commutation relations

$$\begin{aligned} [p_1, p_2] &= [p_2, p_3] = [p_3, p_4] = 0, \\ [p_1, p_3] &= [p_1, p_4] = [p_2, p_4] = 1. \end{aligned} \tag{20}$$

The reason is that since the set is closed under inference, it will necessarily contain the elements $p_1 + p_2$ and $p_3 + p_4$, and the elements $p_1, p_1 + p_2, p_3 + p_4, p_4$ have the commutation relations of Eq. (19). Thus, it must contain a Mermin square.

A similar argument shows that another sufficient condition for a closed under inference set to be contextual is that it contains four elements with the commutation relations

$$\begin{aligned} [p_1, p_2] &= [p_2, p_3] = 0, \\ [p_1, p_3] &= [p_1, p_4] = [p_2, p_4] = [p_3, p_4] = 1. \end{aligned} \tag{21}$$

In this case, since the set is closed under inference, it must also contain the elements $p_1 + p_2$ and $p_2 + p_3$ and the elements $p_1 + p_2, p_2, p_2 + p_3, p_4$ have the commutation relations of Eq. (19).

To determine the possible commutation relations of the elements p_1, \dots, p_{ξ} , we will look at their commutativity graph \mathcal{G} . That is the undirected graph with a vertex for each of p_1, \dots, p_{ξ} and an edge connecting each pair of commuting vertices. Since Ω is noncontextual, the commutation relations of Eq. (19), Eq. (20), and Eq. (21) provide restrictions on the possible commutation relations of the elements p_1, \dots, p_{ξ} of Ω . In terms of the commutativity graph \mathcal{G} , these are forbidden induced subgraphs.¹

The restriction of Eq. (19) says that \mathcal{G} cannot have a four-vertex chordless cycle (C_4) as an induced subgraph and the restriction from Eq. (20) says that \mathcal{G} cannot have a four-vertex path (P_4) as an induced subgraph. These two forbidden induced subgraphs characterize the trivially perfect graphs [40]; i.e., \mathcal{G} must be a trivially perfect graph.

Connected trivially perfect graphs have the the property that they contain a universal vertex [40].² If the commutativity graph \mathcal{G} were connected then there would be an element p_j which commutes with all other elements of $\{p_1, \dots, p_{\xi}\}$. This is also forbidden. Therefore, the graph \mathcal{G} is disconnected.

Given that \mathcal{G} is disconnected, Eq. (21) provides another restriction, namely, that each connected component of \mathcal{G} cannot have a three-vertex path (P_3) as an induced subgraph; i.e., each connected component of \mathcal{G} is a clique. The three forbidden subgraphs are displayed graphically in Fig. 4.

This means we can partition the elements $\{p_1, \dots, p_{\xi}\}$ into disjoint subsets

$$\begin{aligned} \{p_1, \dots, p_{\xi}\} &= \{p_{1,1}, p_{1,2}, \dots, p_{1,\xi_1}\} \\ &\cup \{p_{2,1}, p_{2,2}, \dots, p_{2,\xi_2}\} \\ &\cup \dots \cup \{p_{\pi,1}, p_{\pi,2}, \dots, p_{\pi,\xi_{\pi}}\} \end{aligned} \tag{22}$$

where two elements commute if and only if they are in the same subset in the partition. Since the set $\{p_1, \dots, p_{\xi}\}$ is closed under inference, each subset in the partition must be closed under inference. Now suppose a subset in the partition contained at least two elements. Then since the subset is closed under inference it must also contain their sum. But each of the two elements anticommutes with the elements of all other subsets in the partition, so their sum must commute with the elements of all other subsets in the partition. This is a contradiction. Therefore, each subset in the partition contains a single element. Thus, the elements $\{p_1, \dots, p_{\xi}\}$ of Eq. (18) pairwise anticommute.

Maximal cnc sets are sets of the form Eq. (18) for which ξ is maximal for a given isotropic subspace \tilde{I} . If the isotropic subspace \tilde{I} has dimension $n - m$ where n is the number of qubits and $1 \leq m \leq n$, then the pairwise anticommuting elements p_k which complete the set are elements of the symplectic complement \tilde{I}^{\perp} . This is a m dimensional symplectic subspace, therefore the maximal value of ξ is the largest number of pairwise anticommuting Pauli operators on m qubits. The largest sets of pairwise anticommuting Pauli operators on m qubits have $2m + 1$ elements. This can be seen as follows. Consider the elements $a_k \in E$ given by $T_{a_k} = C_k$, with C_k defined in Eq. (16). The set $\{a_k \mid 1 \leq k \leq 2m\}$ consists of pairwise anticommuting elements. There is an element c , with $T_c = C_{2m+1}$ [cf. Eq. (17)] that anticommutes with each one of the elements in this set. It is the only element in E to do so, since the set of equations

$$[c, a_k] = 1 \quad 1 \leq k \leq 2m \tag{23}$$

has a unique solution. Therefore together with this element we can construct a set of size $2m + 1$.

¹An induced subgraph of a graph is the graph obtained by taking a subset of the vertices of the original graph and all of the edges connecting pairs of vertices in the subset.

²A universal vertex is a vertex that is adjacent to every other vertex in the graph.

We would like to show any other set of pairwise anticommuting elements whose size is $2m$ can be mapped bijectively to the set we constructed. Suppose $\{\tilde{a}_k \mid 1 \leq k \leq 2m\}$ is such a set. By Witt's lemma [41, §20] the function that sends a_k to \tilde{a}_k extends to a linear map $f : E \rightarrow E$ that satisfies $[f(v), f(w)] = [v, w]$ for all $v, w \in E$ (symplectic transformation). Therefore there is a unique element that anticommutes with all the \tilde{a}_k , and it is given by $f(c)$. In particular, $2m + 1$ is the maximal number.

To complete the proof we must show that maximal sets of pairwise anticommuting elements on m qubits with size less than $2m + 1$ do not lead to maximal cnc sets. To see this note that by Witt's lemma, for any maximal anticommuting set of size $2m' + 1$ ($m' < m$), there is a bijection $f : E \rightarrow E$ which maps the set to one of the form Eqs. (16) and (17). Therefore, we can find $m - m'$ independent elements which commute with the set. For example, if $g_1, g_2, \dots, g_{m-m'}$ are the vectors corresponding to Pauli operators $X_{m'+1}, X_{m'+2}, \dots, X_m$, then we could take $f^{-1}(g_1), f^{-1}(g_2), \dots, f^{-1}(g_{m-m'})$. Therefore, the $n - m$ dimensional isotropic subspace can be extended to one with dimension $n - m'$.

This completes the proof. Therefore, all maximal cnc sets have the form Eq. (15), with $\xi = 2m + 1$. ■

A result equivalent to the characterization of Eq. (22) is given in Theorem 3 of Ref. [39].

Tensor products of phase point operators are not typically phase point operators. Consider, for example, two phase point operators with $m = 2$ and $n \geq 2$. Their tensor product does not appear in the classification provided by Theorem 1, as the commutativity graph shows. Physically, such tensor products are not closed under inference, violating Definition 4. Upon closure, they cease to be noncontextual as they then contain a Mermin square. Hence the closures also violate Definition 4.

But there is an exception. If one of the two phase point operators in the tensor product corresponds to an isotropic subspace, i.e., has $m = 0$, then the tensor product is a valid phase point operator. See Appendix D for details.

D. Relation to the stabilizer formalism

The purpose of this section is to describe the relation between positive representability by the quasiprobability distribution W and qubit stabilizer states. We demonstrate that, for all n , the set of positively W -representable states contains the stabilizer mixtures as a strict subset. This is the content of Lemma 4 below. The lemma is based on two examples.

Example 4. Be $|\text{stab}\rangle$ an n -qubit stabilizer state, with isotropic subspace $\tilde{I} \subset E$ corresponding to its stabilizer. Then it is easily verified that \tilde{I} is noncontextual and closed under inference. Namely, \tilde{I} is of form Eq. (15), with $m = 0$, $\xi = 1$.

The next example generalizes Example 1 to n -qubit states.

Example 5. Every n -qubit state of the form $\Psi = \rho_1 \otimes |\text{stab}\rangle_{2,\dots,n}$, with ρ a general one-qubit state and $|\text{stab}\rangle$ an $n - 1$ -qubit pure stabilizer state, is positively representable.

To prove this statement, for any number n of qubits, consider an isotropic subspace $\tilde{I} \subset E$ of rank $n - 1$ representing the stabilizer state $|\text{stab}\rangle_{2,\dots,n}$, and three elements $x, y, z \in E$, such that $T_x = X_1$, $T_y = Y_1$, and $T_z = Z_1$. Define the three

isotropic subspaces $I_x, I_y, I_z \subset E$,

$$I_x = \langle x, \tilde{I} \rangle, \quad I_y = \langle y, \tilde{I} \rangle, \quad I_z = \langle z, \tilde{I} \rangle,$$

and $\Omega_{xyz} := I_x \cup I_y \cup I_z$. Ω_{xyz} is of form Eq. (15), with $m = 1$, $\xi = 3$ and $n \geq 2$, hence cnc by Lemma 3.

We now apply this result to the state $\Psi = \rho_1 \otimes |\text{stab}\rangle_{2,\dots,n}$ above. We can write the constituents as $\rho = \sum_{\gamma_0} W_\rho(\Omega_0, \gamma_0) A_{\Omega_0}^{\gamma_0}$, with $W_\rho \geq 0$ (cf. Example 1), and $|\text{stab}\rangle_{2,\dots,n} = A_{\tilde{I}}^{\tilde{\gamma}}$ (cf. Example 2). We observe that

$$A_{\Omega_0}^{\gamma_0} \otimes A_{\tilde{I}}^{\tilde{\gamma}} = A_{\Omega_{xyz}}^{\gamma},$$

with $\gamma := \gamma_0(a|_1) + \tilde{\gamma}(a|_{2,\dots,n}) \pmod 2$. To see this, recall that $\Omega_0 = \{0, x, y, x\}$, and note that the set Ω_{xyz} can also be written as $\Omega_{xyz} = \tilde{I} \cup (\tilde{I} + x) \cup (\tilde{I} + y) \cup (\tilde{I} + z)$, where the coset $\tilde{I} + x := \{a + x, \forall a \in \tilde{I}\}$, etc. Thus, $\Psi = \sum_{\gamma_0} W_\rho(\Omega_0, \gamma_0) A_{\Omega_{xyz}}^{\gamma}$. Since $W_\rho \geq 0$ by Example 1, the states Ψ are all positively representable. Yet not all these states are mixtures of stabilizer states. Stabilizer mixedness is preserved under partial trace. Now assume that Ψ is a stabilizer mixture for all ρ . Then $\text{Tr}_{2,\dots,n} \Psi = \rho_1$ is also a stabilizer mixture. Contradiction.

We cast the combined conclusion of Examples 4 and 5 as a Lemma.

Lemma 4. For all $n \in \mathbb{N}$, all mixtures of n -qubit stabilizer states are positively representable, and furthermore there exist positively representable states that are not mixtures of stabilizer states.

V. QUANTUM MECHANICAL RULES FOR STATE UPDATE UNDER MEASUREMENT

In the previous sections we have analyzed the generalized phase space \mathcal{V} on which the quasiprobability function W is defined. We now turn to dynamics.

For our setting of QCM this concerns evolution under the free operations, i.e., the Clifford unitaries and Pauli measurements. As already noted in Refs. [19,20], the situation simplifies even further. If the goal is to sample from the joint probability distribution of measurement outcomes—which is the case in quantum computation—then only the update under Pauli measurements needs to be considered.

The Clifford unitaries can be propagated forward in time, thereby conjugating the Pauli measurements into other such measurements, past the final measurement and then discarded. (This redundancy notwithstanding, we will visit the update of W under Clifford unitaries in Sec. VII C, where we prove covariance.) The main results of this section are Theorem 2 and Lemma 5.

Theorem 2. For any $n \in \mathbb{N}$, the set \mathcal{P}_n of positively representable n -qubit quantum states is closed under Pauli measurement.

To describe the dynamics under measurement, we need to set up some further notation. For every set Ω we introduce the derived set $\Omega \times a$. Denoting $\text{Comm}(a) := \{b \in E \mid [a, b] = 0\}$ and $\Omega_a := \Omega \cap \text{Comm}(a)$,

$$\Omega \times a := \Omega_a \cup \{a + b \mid b \in \Omega_a\}, \quad \forall a \notin \Omega. \quad (24)$$

Likewise, we define an update on functions γ invoking the measurement outcome s_a of an observable T_a , namely, $(\cdot) \times s_a : (\gamma : \Omega \rightarrow \mathbb{Z}_2) \mapsto (\gamma \times s_a : \Omega \times a \rightarrow \mathbb{Z}_2)$. We define

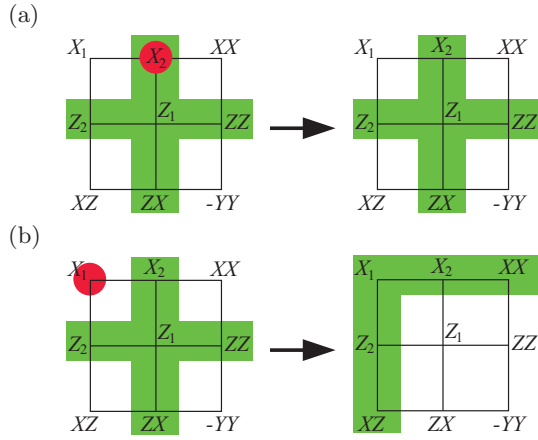


FIG. 5. Update of a cnc set Ω in Mermin's square, under two Pauli measurements. (a) The measured observable X_2 is such that $a(X_2) \in \Omega$; hence the update proceeds by Eq. (26a). (b) The measured observable X_1 is such that $a(X_1) \notin \Omega$; hence the update proceeds by Eq. (26b).

this update only for $(\Omega, \gamma) \in \mathcal{V}$, and only for $a \notin \Omega$.³ The updated function $\gamma \times s_a : \Omega \times a \rightarrow \mathbb{Z}_2$ is given by

$$\gamma \times s_a(b) := \gamma(b), \quad \forall b \in \Omega_a, \quad (25a)$$

$$\gamma \times s_a(b) := \gamma(a+b) + s_a + \beta(a, b), \quad \forall a+b \in \Omega_a. \quad (25b)$$

The rules of Eq. (25) are used to formulate the update rule for phase point operators of Eq. (2) under Pauli measurement.

Remark. Update rules similar to Eq. (25) have been used previously [42] to construct a ψ -epistemic model of the multiqubit stabilizer formalism. Those rules update the value assignments in the same way but are applied under different conditions. Specifically, the update in Ref. [42] does not refer to general sets Ω satisfying the conditions of Def 4.

Lemma 5. Denote the projectors $P_a(s_a) := (I + (-1)^{s_a} T_a)/2$, and let A_Ω^γ a phase point operator defined through Eq. (2), with $(\Omega, \gamma) \in \mathcal{V}$ satisfying the conditions of Definition 4. Then the effect of a measurement of the Pauli observable T_a with outcome s_a on A_Ω^γ is

$$P_a(s_a) A_\Omega^\gamma P_a(s_a) = \delta_{s_a, \gamma(a)} \frac{A_\Omega^\gamma + A_\Omega^{\gamma+[a, \cdot]}}{2}, \quad \text{if } a \in \Omega, \quad (26a)$$

$$P_a(s_a) A_\Omega^\gamma P_a(s_a) = \frac{1}{2} A_{\Omega \times a}^{\gamma \times s_a}, \quad \text{if } a \notin \Omega. \quad (26b)$$

Example 2, continued. Equation (26) entails the update of both the sets Ω and the functions γ . Here we consider only the former. Figure 5 displays the update of the set Ω shown in Fig. 1(a), under the measurement of (a) the observable X_2 , with $a(X_2) \in \Omega$, and (b) the observable X_1 , with $a(X_1) \notin \Omega$.

³The definitions of $\Omega \times a$ and $\gamma \times s_a$ can without modification be extended to $a \in \Omega$. However, in that case the function values $\gamma \times s_a(b)$ can be determined through both Eqs. (25a) and (25b), and we need to check consistency. These inferences are indeed consistent, as a consequence of Eq. (10). Since we do not need the case of $a \in \Omega$ subsequently, we skip the details of the argument.

In preparation for the proof of Lemma 5 it is useful to state two relations of the function β for $d = 2$. With the definition Eq. (5) of β and Eq. (3), the operator identities $T_a T_a = I$ and $T_b I = T_b$ imply that

$$\beta(a, a) = \beta(a, 0) = \gamma(0), \quad \forall a \in E. \quad (27)$$

Furthermore, evaluating $d\beta(a, a, 0) = 0$ (see Eqs. (7) and (11)), and using Eq. (27) yields

$$\beta(a, b) = \beta(a, a+b), \quad \forall a, b. \quad (28)$$

To prove Lemma 5 we also need the following result.

Lemma 6. If $\Omega \subset E$ is noncontextual and closed under inference, then so is Ω_a , for all $a \in E$.

Proof of Lemma 6. First consider closure. Assume that $c, d \in \Omega_a$ and $[c, d] = 0$. Then $c, d \in \Omega$, and $c+d \in \Omega$, since Ω is closed by assumption. Further, $[c, a] = [d, a] = 0$ implies $[c+d, a] = 0$, and hence $c+d \in \Omega_a$. Ω_a is thus closed.

Now consider noncontextuality. Since Ω is noncontextual, there exists a function γ such that $d\gamma = \beta$ on Ω . Since Ω_a is closed, β can be properly restricted to $\mathcal{C}(\Omega_a)$, and so can γ . Hence, $d\gamma|_{\mathcal{C}(\Omega_a)} = \beta|_{\mathcal{C}(\Omega_a)}$. Thus, Ω_a is noncontextual. ■

Proof of Lemma 5. Under the measurement of T_a with outcome $s_a \in \mathbb{Z}_2$ we have

$$\begin{aligned} & \frac{I + (-1)^{s_a} T_a}{2} A_\Omega^\gamma \frac{I + (-1)^{s_a} T_a}{2} \\ &= \frac{I + (-1)^{s_a} T_a}{2} \frac{1}{2^n} \sum_{b \in \Omega_a} (-1)^{\gamma(b)} T_b \\ &= \frac{1}{2 \times 2^n} \sum_{b \in \Omega_a} (-1)^{\gamma(b)} T_b + \frac{(-1)^{s_a}}{2 \times 2^n} \sum_{b \in \Omega_a} (-1)^{\gamma(b)} T_a T_b. \end{aligned} \quad (29)$$

From hereon we need to distinguish two cases, $a \in \Omega$ and $a \notin \Omega$.

Case I: $a \in \Omega$. Focusing on the second term in the expansion Eq. (29),

$$\begin{aligned} (-1)^{s_a} \sum_{b \in \Omega_a} (-1)^{\gamma(b)} T_a T_b &= (-1)^{s_a} \sum_{b \in \Omega_a} (-1)^{\gamma(b) + \beta(a, b)} T_{a+b} \\ &= (-1)^{s_a + \gamma(a)} \sum_{b \in \Omega_a} (-1)^{\gamma(a+b)} T_{a+b} \\ &= (-1)^{s_a + \gamma(a)} \sum_{a+b \in \Omega_a} (-1)^{\gamma(a+b)} T_{a+b} \\ &= (-1)^{s_a + \gamma(a)} \sum_{b \in \Omega_a} (-1)^{\gamma(b)} T_b. \end{aligned}$$

Therein, in the first line we have used Eq. (5), in the second line Eq. (10), in the third line the completeness of Ω_a under inference (Lemma 6), and the fourth line is just a relabeling of the elements in Ω_a . Inserting this result in the above expansion Eq. (29), we find

$$P_a(s_a) A_\Omega^\gamma P_a(s_a) = \delta_{s_a, \gamma(a)} \frac{1}{2^n} \sum_{b \in \Omega_a} (-1)^{\gamma(b)} T_b, \quad (30)$$

and Eq. (26a) follows.

Case II: $a \notin \Omega$. Substituting $b \rightarrow a+b$ in Eq. (25b) gives $\gamma \times s_a(a+b) = \gamma(b) + s_a + \beta(a, a+b)$, for $b \in \Omega_a$. With Eq. (28) we obtain

$$\gamma \times s_a(a+b) = \gamma(b) + s_a + \beta(a, b), \quad \forall b \in \Omega_a. \quad (31)$$

With this, we now look at the second term in the expansion Eq. (29),

$$\begin{aligned} (-1)^{s_a} \sum_{b \in \Omega_a} (-1)^{\gamma(b)} T_a T_b &= (-1)^{s_a} \sum_{b \in \Omega_a} (-1)^{\gamma(b) + \beta(a,b)} T_{a+b} \\ &= \sum_{b \in \Omega_a} (-1)^{\gamma \times s_a(a+b)} T_{a+b}. \end{aligned}$$

The first line above follows with Eq. (5), and the second with Eq. (31).

Considering the first term in the expansion Eq. (29), with Eq. (25a) we have

$$\sum_{b \in \Omega_a} (-1)^{\gamma(b)} T_b = \sum_{b \in \Omega_a} (-1)^{\gamma \times s_a(b)} T_b.$$

Inserting the above expressions for the two terms in Eq. (29), and using the definition Eq. (24) of $\Omega \times a$, we obtain Eq. (26b). ■

We have so far shown how the phase point operators can be updated under measurement once. We still need to show that this update can be iterated. This requires that the phase point operators appearing on the r.h.s. of Eq. (26) satisfy the consistency constraints of Definition 4.

Lemma 7. If $(\Omega, \gamma) \in \mathcal{V}$ then $(\Omega, \gamma + [a, \cdot]) \in \mathcal{V}$, for all $a \in \Omega$, and $(\Omega \times a, \gamma \times s_a) \in \mathcal{V}$, for all $a \notin \Omega$ and $s_a \in \mathbb{Z}_2$.

The proof of Lemma 7 is given in Appendix B.

Proof of Theorem 2. Consider a state $\rho \in \mathcal{P}_n$, and a measurement of the Pauli observable T_a on it. Assume that the measurement outcome s_a can occur, $p_a(s_a) := \text{Tr}(P_a(s_a)\rho) > 0$. We have to show that under these conditions, the postmeasurement state

$$\rho' = \frac{P_a(s_a)\rho P_a(s_a)}{p_a(s_a)}$$

is also contained in the set \mathcal{P}_n .

Denote $\bar{\delta}_{a \in \Omega} := 1 - \delta_{a \in \Omega}$. Then, with Lemma 5 and the state expansion Eq. (4) of ρ , we have

$$\begin{aligned} \rho' &= \sum_{(\Omega, \gamma) \in \mathcal{V}} \frac{W_\rho(\Omega, \gamma)}{p_a(s_a)} \left(\delta_{a \in \Omega} \delta_{s_a, \gamma(a)} \frac{A_\Omega^\gamma + A_\Omega^{\gamma + [a, \cdot]}}{2} \right. \\ &\quad \left. + \frac{1}{2} \bar{\delta}_{a \in \Omega} A_{\Omega \times a}^{\gamma \times s_a} \right). \end{aligned} \quad (32)$$

Thus, ρ' can be represented by a quasiprobability distribution $W_{\rho'}$ with elements

$$\begin{aligned} W_{\rho'}(\Omega', \gamma') &= \sum_{(\Omega, \gamma) \in \mathcal{V}} \frac{W_\rho(\Omega, \gamma)}{2p_a(s_a)} \left[\delta_{a \in \Omega} \delta_{s_a, \gamma(a)} (\delta_{(\Omega', \gamma'), (\Omega, \gamma)} \right. \\ &\quad \left. + \delta_{(\Omega', \gamma'), (\Omega, \gamma + [a, \cdot])}) + \bar{\delta}_{a \in \Omega} \delta_{(\Omega', \gamma'), (\Omega \times a, \gamma \times s_a)} \right]. \end{aligned} \quad (33)$$

TABLE I. Classical simulation algorithm for sampling from the joint probability distribution of a sequence of Pauli measurements on a positively represented initial quantum state.

1. Draw a sample $(\Omega, \gamma) \in \mathcal{V}$ according to the probability distribution W_ρ representing the initial quantum state ρ .
2. For the observables $T_{a_1}, T_{a_2}, \dots, T_{a_N}$ measured in this sequence, repeat the following steps.
For the i th measurement, set $a := a_i$.
If $a \in \Omega$ then Ω is unchanged. Output the value $s_a = \gamma(a)$.
Flip a coin.
if “heads” then $\gamma \rightarrow \gamma$,
if “tails” then $\gamma \rightarrow \gamma + [a, \cdot]$.
If $a \notin \Omega$ then $\Omega \rightarrow \Omega \times a$. Flip a coin.
if “heads” then $s_a = 0$,
if “tails” then $s_a = 1$.
Output this value s_a . Update $\gamma \rightarrow \gamma \times s_a$, through Eq. (25).

The $W_{\rho'}(\Omega', \gamma')$ are thus linear combinations of $W_\rho(\Omega, \gamma)$ with non-negative coefficients [0 or $1/2p_a(s_a)$]. Since the $W_\rho(\Omega, \gamma)$ are non-negative by assumption, it follows that $W_{\rho'}(\Omega', \gamma') \geq 0$, for all $(\Omega', \gamma') \in \mathcal{V}$. ■

VI. CLASSICAL SIMULATION FOR $W_\rho \geq 0$

A. Simulation algorithm

We now turn to the question of how hard it is to classically simulate the outcome statistics for a sequence of Pauli measurements on an initial quantum state. In this regard, we show that if the initial quantum state is positively represented and the corresponding probability distribution W can be efficiently sampled from, then the statistics of the measurement outcomes can be efficiently simulated.

The classical simulation procedure in Table I describes weak simulation [24,25]; i.e., it outputs one sample from the joint probability distribution $p(s_{a_1}, s_{a_2}, \dots, s_{a_N})$ of outcomes corresponding to a sequence of measurements of Pauli operators $T_{a_1}, T_{a_2}, \dots, T_{a_N}$ (T_{a_1} is measured first, T_{a_N} last). If more than one sample are desired, the procedure is just repeated. We note that the measurement can be adaptive; i.e., it is not necessary for the simulation algorithm that a measurement sequence is committed to at the beginning. As a special case of this, the measured observables may depend on earlier measurement outcomes.

We have the following result.

Theorem 3. If for an initial quantum state ρ it holds that $W_\rho \geq 0$ and furthermore W_ρ can be efficiently sampled from, then the output distribution of all sequences of Pauli measurements, possibly interspersed with Clifford gates, on ρ can be classically efficiently sampled from.

As a first application of Theorem 3, we return to Example 2, Mermin’s square.

Example 2 continued. How much memory capacity is needed to classically simulate measurements of the observables in Mermin’s square? We first turn to the state-independent case, which was previously discussed in Ref. [34]. The task is to devise a classical algorithm that outputs an outcome sequence for any given sequence of Pauli measurements, which can occur according to quantum

TABLE II. Number of points in phase space as a function of $\{m\}$.

m	0	1	$\{1, 2\}$
2 rebits	24	72	120
2 qubits	60	240	432

mechanics. The measurement sequence can be of any length, and the measurements therein may be commuting or anticommuting. In Ref. [34], a lower bound on the memory cost of any such simulation was established, $\log_2 24$ bits; and a specific model was constructed that attains it.

The classical simulation algorithm of Table I also saturates this limit. To show this, we use as cnc sets Ω the six maximal isotropic subspaces of two rebits; cf. Fig. 1(b). This set of sets Ω is closed under update by Pauli measurement, as described by Eq. (26). For each such set Ω , each value assignment γ is specified by two evaluations [the other evaluations then follow via Eq. (10)]. There are thus four functions γ for each cnc set Ω , hence 24 combinations in total, which is the same as in Ref. [34].

We now turn to the state-dependent version of the problem. How much memory is needed to sample from the correct outcome statistics for arbitrary measurement sequences, for any two-rebit state ρ with $W_\rho \geq 0$, and given the capability to sample from W_ρ ? This problem is harder than the former: Not only must the sequence of outcomes be internally consistent for all measurement sequences, but also it needs to represent the state ρ .

Memory cost now depends on the state ρ . If ρ is a mixture of stabilizer states, i.e., the sets Ω can be limited to $m = 0$, then the classical simulation algorithm of Table I can still run on $\log_2 24 \approx 4.59$ bits.

If sets Ω with $m = 1$ are included in the expansion, then more two-rebit states ρ can be positively represented (among them, for example, $|T\rangle_1 \otimes |T\rangle_2$), but on the other hand, memory consumption goes up. For $m = 1$, there are $3^2 \times 2^3$ pairs (Ω, γ) ; cf. Fig. 1(a). Hence the memory consumption for configurations with $m = 1$ is $\log_2 72 \approx 6.17$ bits. (Note that the sets Ω for $m = 0$ are not maximal. If sets with $m = 1$ are included, then sets with $m = 0$ can be omitted without loss.) The size $|\mathcal{V}(\{m\})|$ of the phase space versus the maximum value of m is displayed in Table II. The memory cost is $\log_2 |\mathcal{V}(\{m\})|$. The volume fraction of positively representable two-rebit and two-qubit states is displayed in Table III, for various sets $\{m\}$.

B. Correctness and efficiency of the classical simulation

In preparation for the proof of correctness of the classical simulation algorithm, we introduce the following notation. Given a probability distribution W_ρ , there are two objects that the classical simulation algorithm needs to reproduce correctly, namely, the probability $p_a(s_a)$ for the outcomes $s_a \in \mathbb{Z}_2$ of the measurement of any Pauli observable T_a , and the postmeasurement state ρ' . There are two ways of obtaining these quantities, a quantum-mechanical one and a classical one using the simulation algorithm of Sec. VIA.

TABLE III. Volume fraction of state space filled by the positively representable states, as a function of $\{m\}$; (top) two rebits, (bottom) two qubits. The volume fraction V_+/V was obtained numerically, by sampling 10^6 random states according to the Fubini-Study measure for pure states (second row) and the Hilbert-Schmidt measure for mixed states (third row). The first column, $m = 0$, describes mixtures of stabilizer states, and the last column hyperoctahedral states [38] for comparison.

Two rebits				
m	0	1	$\{1, 2\}$	Hyperoctahedral
V_+/V [pure]	0	1	1	0
V_+/V [mixed]	0.144	1	1	0.924
Two qubits				
m	0	1	$\{1, 2\}$	Hyperoctahedral
V_+/V [pure]	0	0.980	0.980	0
V_+/V [mixed]	0.009	1	1	0.568

Regarding the outcome probability $p_a(s_a)$ given W_ρ , the quantum mechanical way first obtains the corresponding quantum state ρ from W_ρ through Eq. (4). This is represented by a map $R : W_\rho \mapsto \rho$. Second, from ρ , the outcome probability $p_a(s_a)$ is obtained via the Born rule, $p_a(s_a) = \text{Tr}(P_a(s_a)\rho)$. This is represented by a map $\pi_q(a) : \rho \mapsto p_a$. The classical way uses the algorithm of Sec. VIA to obtain p_a . This is represented by a map $\pi_c(a) : W_\rho \mapsto p_a$.

Likewise, the quantum mechanical way of obtaining the postmeasurement state ρ' from W_ρ proceeds by first applying the map R [Eq. (4)] to obtain ρ , and second by obtaining ρ' from ρ through the Dirac projection postulate. The second step is represented by a map $\Pi_q(a)$.

The classical way of obtaining ρ' from W_ρ proceeds by first using the simulation algorithm to obtain $W_{\rho'}$, and second by mapping $W_{\rho'}$ to ρ' using the map R . The first step in this procedure is represented by the map $\Pi_c(a)$.

The classical simulation algorithm of Sec. VIA is correct only if the quantum and the classical ways of computing $p_a(s_a)$ and ρ' agree. That is, we require the diagrams in Fig. 6 to commute.

Lemma 8. The diagrams of Fig. 6 commute.

Proof of Lemma 8. We discuss the outcome probability and the postmeasurement state separately.

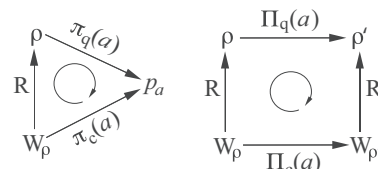


FIG. 6. Diagrams representing the quantum and the classical way of calculating the probability of measurement outcomes (left) and the postmeasurement state (right).

Outcome probability $p_a(s_a)$. Then the quantum mechanical expression for $p_a(s_a)$ is

$$\begin{aligned} p_a(s_a) &= \sum_{(\Omega, \gamma) \in \mathcal{V}} W_\rho(\Omega, \gamma) \text{Tr} \left[\frac{I + (-1)^{s_a} T_a}{2} A_\Omega^\gamma \right] \\ &= \sum_{(\Omega, \gamma) \in \mathcal{V}} W_\rho(\Omega, \gamma) \left(\delta_{a \in \Omega} \delta_{s_a, \gamma(a)} + \frac{1}{2} \bar{\delta}_{a \in \Omega} \right). \end{aligned} \quad (34)$$

The classical expression $p_a^{(c)}(s_a)$ for $p_a(s_a)$ obtained through the algorithm of Sec. VI A is as follows. If $a \in \Omega$, then the conditional probability for the outcome s_a given the state (Ω, γ) is $\delta_{s_a, \gamma(a)}$. If $a \notin \Omega$, then the conditional probability for the outcome s_a is $1/2$. Thus,

$$p_a^{(c)}(s_a) = \sum_{(\Omega, \gamma) \in \mathcal{V}} W_\rho(\Omega, \gamma) \left(\delta_{a \in \Omega} \delta_{s_a, \gamma(a)} + \frac{1}{2} \bar{\delta}_{a \in \Omega} \right).$$

By comparing the two expressions, we find that $p_a(s_a) = p_a^{(c)}(s_a)$ for all a, s_a , and the left diagram of Eq. (6) thus commutes.

Postmeasurement state ρ' . The quantum mechanical expression for the postmeasurement state ρ' has already been given in Eq. (32), and we now derive the corresponding expression $\rho'_{(c)}$ that follows from the classical simulation algorithm.

We consider the joint probability $p((\Omega', \gamma') \cap s_a)$ of obtaining the outcome s_a in the measurement of T_a and ending up in the state (Ω', γ') . We may invoke conditional probabilities in two ways:

$$\begin{aligned} p((\Omega', \gamma') \cap s_a) &= \sum_{(\Omega, \gamma) \in \mathcal{V}} p((\Omega', \gamma') \cap s_a | (\Omega, \gamma)) W_\rho(\Omega, \gamma) \\ &= p((\Omega', \gamma') | s_a) p_a(s_a). \end{aligned}$$

Noting that $p((\Omega', \gamma') | s_a) = W_{\rho'}(\Omega', \gamma')$, and equating the two above expressions we find

$$W_{\rho'}(\Omega', \gamma') = \sum_{(\Omega, \gamma) \in \mathcal{V}} \frac{W_\rho(\Omega, \gamma)}{p_a(s_a)} p((\Omega', \gamma') \cap s_a | (\Omega, \gamma)). \quad (35)$$

We now infer the conditional probabilities $p((\Omega', \gamma') \cap s_a | (\Omega, \gamma))$ from the classical simulation algorithm of Sec. VI A,

$$\begin{aligned} p((\Omega', \gamma') \cap s_a | (\Omega, \gamma)) &= \begin{cases} \frac{1}{2} \delta_{s_a, \gamma(a)} \delta_{\Omega', \Omega} (\delta_{\gamma', \gamma} + \delta_{\gamma', \gamma + [a, \cdot]}), & \text{if } a \in \Omega, \\ \frac{1}{2} \delta_{\Omega', \Omega \times a} \delta_{\gamma', \gamma \times s_a}, & \text{if } a \notin \Omega. \end{cases} \end{aligned}$$

Inserting this into Eq. (35), and using the resulting expression in Eq. (4), i.e., applying the map R , we obtain

$$\begin{aligned} \rho'_{(c)} &= \sum_{(\Omega', \gamma') \in \mathcal{V}} \sum_{(\Omega, \gamma) \in \mathcal{V}} \frac{W_\rho(\Omega, \gamma)}{2 p_a(s_a)} \left[\delta_{a \in \Omega} \delta_{s_a, \gamma(a)} \delta_{\Omega', \Omega} (\delta_{\gamma', \gamma} \right. \\ &\quad \left. + \delta_{\gamma', \gamma + [a, \cdot]}) + \bar{\delta}_{a \in \Omega} \delta_{\Omega', \Omega \times a} \delta_{\gamma', \gamma \times s_a} \right] A_{\Omega'}^{\gamma'} \\ &= \sum_{(\Omega, \gamma) \in \mathcal{V}} \frac{W_\rho(\Omega, \gamma)}{2 p_a(s_a)} \left[\delta_{a \in \Omega} \delta_{s_a, \gamma(a)} (A_\Omega^\gamma + A_\Omega^{\gamma + [a, \cdot]}) \right. \\ &\quad \left. + \bar{\delta}_{a \in \Omega} A_{\Omega \times a}^{\gamma \times s_a} \right]. \end{aligned}$$

Comparing the last expression with Eq. (32), we find that $\rho'_{(c)} = \rho'$ for all a, s_a , and the right diagram in Fig. 6 thus commutes. ■

Proof of Theorem 3. As explained in Sec. V, we only need to discuss sequences of Pauli measurements. For those, we show that the algorithm of Table I is correct, and, if the initial W_ρ can be efficiently sampled from, it is also computationally efficient. (i) *Correctness.* Denote by $\rho(t)$ the state before the t th measurement. With Lemma 8, by induction on the right diagram in Eq. (6), if $W_{\rho(1)}$ represents the initial state $\rho(1)$, then $W_{\rho(t)}$ represents $\rho(t)$ for all time steps $t = 1, \dots, N$. Then, by the left diagram in Eq. (6), the outcome probabilities $p_{a_t}(s_{a_t} | \mathbf{s}_{<t})$, with $\mathbf{s}_{<t} = (s_{a_1}, \dots, s_{a_{t-1}})$ the measurement record prior to time t , are also correct. Thus the joint outcome probability sampled from

$$p_{a_1, \dots, a_N}(s_{a_1}, \dots, s_{a_N}) = \prod_{t=1}^N p_{a_t}(s_{a_t} | \mathbf{s}_{<t})$$

is also correct.

(ii) *Efficiency.* We recall that all cnc sets Ω are unions of $O(n)$ isotropic spaces Ω_i (Theorem 1). Further, each Ω_i defines a stabilizer group

$$T_{\Omega_i}^\gamma := \{T_a^\gamma := (-1)^{\gamma(a)} T_a, a \in \Omega_i\}. \quad (36)$$

This allows us to describe $(\Omega, \gamma) \in \mathcal{V}$ using polynomial memory by storing $O(n)$ stabilizer tables of size $O(n^2)$ [17,43]. Indeed, by Definitions 2 and 3 and Lemma 3, $T_{\Omega_i}^\gamma$ is a closed commutative group. Furthermore, with Definition 3, it holds that $T_a^\gamma T_b^\gamma = T_{a+b}^\gamma, \forall a, b \in \Omega_i$. This implies the existence of a nontrivial stabilized subspace: $P_{\Omega_i}^\gamma := \sum_{a \in \Omega_i} T_a^\gamma / |\Omega_i|$ is a common +1-eigenprojector of every $T_a \in T_{\Omega_i}^\gamma$ as $T_a^\gamma P_{\Omega_i}^\gamma = \sum_{b \in \Omega_i} \frac{T_{a+b}^\gamma}{|\Omega_i|} = \sum_{b' \in \Omega_i} \frac{T_{b'}^\gamma}{|\Omega_i|} = P_{\Omega_i}^\gamma, \forall T_{\Omega_i}^\gamma$, which also implies $P_{\Omega_i}^{\gamma^2} = P_{\Omega_i}^\gamma$.

We now note that the update rules in algorithm I, namely, (i) checking whether $a \in \Omega$, (ii) evaluating γ on $a \in \Omega$, (iii) updating $\gamma \rightarrow \gamma + [a, \cdot]$, (iv) $\Omega \rightarrow \Omega \times a$, and (v) $\gamma \rightarrow \gamma \times s_a$, implement tasks that admit efficient classical algorithms in the stabilizer formalism [17,43]. *Rules (i) and (ii):* To test $a \in \Omega$, we check whether $a \in \Omega_i, i = 1, \dots, O(n)$. If $a \in \Omega_j$ for some value of j , then $\gamma(a)$ is computed as the bit determining the phase of the stabilizer operator $T_a^\gamma \in T_{\Omega_j}^\gamma$. Both tasks can be solved classically efficiently via Gaussian elimination given the stabilizer table data [17,43]. *Rule (iii):* γ is updated to $\gamma' = \gamma + [a, \cdot]$ by (classically efficiently) evaluating $\gamma(\cdot) + [a, \cdot]$ on the generators of every Ω_i . *Rules (iv) and (v):* For all $j, T_{\Omega_j \times a}^{\gamma | \Omega_j \times s_a}$ is the stabilizer group resulting from the measurement of T_a with outcome s_a on a state with stabilizer group $T_{\Omega_j}^{\gamma | \Omega_j}$. This update can be efficiently performed using the standard measurement update rule of Refs. [17,43] to every stabilizer table in the description of (Ω, γ) . Thus, all steps of the algorithm run in polynomial time. ■

VII. THE CASE OF $W_\rho < 0$

As we have established in the previous sections, $W_\rho < 0$ is a precondition for quantum speedup. When the initial state is

represented by a quasiprobability rather than a true probability function, a standard problem of interest is estimating outcome probabilities for sequences of measurements. An established method for probability estimation is Ref. [28], utilizing the Hoeffding bound. Note that probability estimation is a different problem than weak simulation [24] and is not efficiently adaptive.

A. Robustness

In close analogy to the “robustness of magic” [27] \mathfrak{R}_S (the subscript S is for “stabilizer”), we define a phase space robustness \mathfrak{R} , through

$$\mathfrak{R}(\rho) := \min_{W | \langle \mathcal{A}, W \rangle = \rho} \|W\|_1, \quad (37)$$

with $\langle \mathcal{A}, W \rangle := \sum_{\alpha \in \mathcal{V}} W_{\alpha} A_{\alpha}$.

Since the definitions of the robustness \mathfrak{R} and of the robustness of magic \mathfrak{R}_S [27] are so similar, one may wonder if there is a relation between them. This is indeed the case; namely, we have the following result.

Lemma 9. For all quantum states ρ , of any number n of qubits, the phase space robustness $\mathfrak{R}(\rho)$ and the robustness of magic $\mathfrak{R}_S(\rho)$ are related via

$$\mathfrak{R}(\rho) \leq \mathfrak{R}_S(\rho) \leq (4n + 1) \mathfrak{R}(\rho). \quad (38)$$

Thus, the phase space robustness \mathfrak{R} is never larger than the robustness of magic, but can only be moderately smaller. The proof of Lemma 9 is given in Appendix C.

B. Hardness of classical simulation

The Hoeffding bound says that the number N of samples required to estimate the output probability distribution up to an error ϵ scales as $N \sim \mathcal{M}^2/\epsilon^2$, where \mathcal{M} is a measure of the negativity contained in the quantum process. In our case, the operations are positivity-preserving, and all negativity comes from the initial state. The algorithm of Pashayan *et al.* [28], when applied to our setting, says that the number N of samples required to estimate the output probability scales as

$$N \sim \frac{\mathfrak{R}(\rho_{\text{init}})^2}{\epsilon^2}.$$

Thus, the robustness $\mathfrak{R}(\rho_{\text{init}})$ of the initial state ρ_{init} is the critical parameter determining the classical hardness of probability estimation.

The same relation, with the robustness \mathfrak{R} replaced by the robustness of magic \mathfrak{R}_S holds for the classical simulation based on quasiprobability distributions over stabilizer states [27]. Lemma 9 above is therefore of interest for relating the operational costs of the two simulation methods.

Classical simulation also requires a quasiprobability function $W_{\mu^{\otimes n}}$ for n copies of the magic state μ . Since the n -qubit phase space is large, the numerical optimization to obtain the least-negative expansion $W_{\mu^{\otimes n}}$ is computationally costly. However, we can apply a similar splitting into smaller blocks of magic states as in the stabilizer case [27]. The computational cost for providing the expansion is then a function of block size rather than total number of copies n . The one-norm of the resulting expansion is smaller than of the stabilizer

expansion, by a factor that is constant in n . Details are given in Appendix D.

C. Elements of a resource theory based on W

It is illuminating to discuss QCM within the framework of resource theories. Every resource theory has three main operational components [44]: (i) the resource(s), (ii) the non-resources, or free states, and (iii) the free operations.

In the physical setting of our interest, the resources are quantum states which cannot be positively represented by W (cf. Theorem 3). The free operations are Clifford unitaries and Pauli measurements. The free states are those that can be created from the free operations from a completely mixed state, i.e., all mixtures of stabilizer states.

We observe that there is a third class of states which are neither resources nor free, namely, the positively representable states which are not mixtures of stabilizer states. Such states are called (iv) bound magic states. We have seen an example of them in Sec. IV D, the general one-qubit states tensored with a stabilizer state on arbitrarily many qubits.

The reason for calling those states “bound magic” is that they cannot be distilled into computationally useful ones by free operations. In our setting, by Theorem 2, positive representability is an invariant under the free operations. Hence, bound states can be converted only into other bound states or into free states by the free operations, but never into a resource.

The question of interconvertibility may more generally be asked for resource states. To facilitate this discussion, one may identify monotones, i.e., real-valued functions on the state space that never increase under the free operations. The main result of this section is that the robustness \mathfrak{R} , defined in Eq. (37) and already known to measure hardness of classical simulation by sampling, is a monotone.

Theorem 4. The robustness \mathfrak{R} is a monotone under all Clifford unitaries and Pauli measurements.

As part of the proof of Theorem 4, we now discuss an important structural property of the quasiprobability function W , namely, its covariance under Clifford unitaries. Let Cl_n be the n -qubit Clifford group. It acts on the n -qubit Pauli operators via

$$h(T_a) := hT_a h^\dagger = (-1)^{\Phi_h(a)} T_{ha}, \quad \forall h \in \text{Cl}_n.$$

This relation simultaneously defines the phase function Φ and the action of Cl_n on E . It implies an action of the Clifford group on the phase point operators A_{Ω}^{γ} , which in turn induces an action on the sets Ω and the functions γ , via

$$h(A_{\Omega}^{\gamma}) = \frac{1}{2^n} \sum_{a \in \Omega} (-1)^{\gamma(a)} h(T_a) = \frac{1}{2^n} \sum_{b \in \Omega'} (-1)^{\gamma'(b)} T_b.$$

Therein, the set Ω' is defined as $\Omega' := \{ha, a \in \Omega\}$, and the function $\gamma' : \Omega' \rightarrow \mathbb{Z}_2$ is given by

$$\gamma'(ha) := \gamma(a) + \Phi_h(a), \quad \forall a \in \Omega.$$

Henceforth we denote Ω' as $h \cdot \Omega$ and γ' as $h \cdot \gamma$, to emphasize the dependence on $h \in \text{Cl}_n$.

For use in the proof below we quote Lemma 3 from Ref. [35], which says that, for any face $(a, b) \in \Omega \times \Omega$,

$$\Phi_h(\partial(a, b)) = \beta(ha, hb) + \beta(a, b) \pmod{2}.$$

We then have the following result.

Lemma 10. \mathcal{V} is mapped to itself under Cl_n , and the quasiprobability function W transforms covariantly. That is, if the state ρ can be described by W_ρ through Eq. (4), then for any $h \in \text{Cl}_n$ the state $h\rho h^\dagger$ can be described by a quasiprobability function $W_{h\rho h^\dagger}$ defined by

$$W_{h\rho h^\dagger}(\Omega, \gamma) := W_\rho(h^{-1} \cdot \Omega, h^{-1} \cdot \gamma).$$

Remark 3. We say “the state ρ can be described by W_ρ ” rather than “is described” because W_ρ is not unique.

Proof of Lemma 10. First, we show that the phase space \mathcal{V} is closed under the action of Cl_n , i.e., if $(\Omega, \gamma) \in \mathcal{V}$ then $(\Omega', \gamma') \in \mathcal{V}$. The four items in Definition 4 need to be checked. (i) *Closedness under inference.* Assume that $c, d \in \Omega'$, and $[c, d] = 0$. Then there exist $a, b \in \Omega$ such that $c = ha$, $d = hb$ and $[a, b] = 0$. Then $c + d = ha + hb = h(a + b) \in \Omega'$, since $a + b \in \Omega$ by the assumption of closedness. Hence Ω' is closed under inference.

(iii) γ' satisfies Eq. (10). With the definition of γ' we have (all addition mod 2)

$$\begin{aligned} d\gamma'(ha, hb) &= d\gamma(a, b) + \Phi_h(\partial(a, b)) \\ &= d\gamma(a, b) + \beta(ha, hb) + \beta(a, b) \\ &= \beta(ha, hb). \end{aligned}$$

Therein, in the second line we have used Eq. (10). Thus, γ' satisfies Eq. (10) on its domain.

(ii) Ω' is noncontextual. With γ' we have just proved the existence of a function on Ω' that satisfies Eq. (10).

(iv) γ' satisfies Eq. (3). Since γ satisfies Eq. (3), it follows $I = h(I) = h[(-1)^{\gamma(0)} T_0] = (-1)^{\gamma(0) + \Phi_h(0)} T_0 = (-1)^{\gamma'(0)} T_0$. Equation (3) is thus satisfied for γ' .

Hence, if $(\Omega, \gamma) \in \mathcal{V}$ then $(\Omega', \gamma') \in \mathcal{V}$, as claimed.

Next we turn to the covariance of W under Cl_n . We have

$$\begin{aligned} h\rho h^\dagger &= \sum_{(\Omega, \gamma) \in \mathcal{V}} W_\rho(\Omega, \gamma) h(A_\Omega^\gamma) \\ &= \sum_{(\Omega, \gamma) \in \mathcal{V}} W_\rho(\Omega, \gamma) A_{h\Omega}^{h\gamma} \\ &= \sum_{(\Omega, \gamma) \in \mathcal{V}} W_\rho(h^{-1} \cdot \Omega, h^{-1} \cdot \gamma) A_\Omega^\gamma. \end{aligned}$$

Comparing the last expression with the expansion Eq. (4) for $h\rho h^\dagger$, we find that for all $h \in \text{Cl}_n$, the quasiprobability distribution $W_{h\rho h^\dagger}$ defined by

$$W_{h\rho h^\dagger}(\Omega, \gamma) = W_\rho(h^{-1} \cdot \Omega, h^{-1} \cdot \gamma) \quad (39)$$

describes the state $h\rho h^\dagger$. This is the covariance condition. ■

We are now ready to prove the monotonicity of \mathfrak{R} , as stated in Theorem 4.

Proof of Theorem 4. (a) Clifford unitaries. With Lemma 10, we have that for any n -qubit Clifford gate h applied to any n -qubit state ρ , the quasiprobability distribution $W_{h\rho h^\dagger}$ can be related to W_ρ via the covariance condition Eq. (39). Since W is

nonunique, there may a priori be a representation $W'_{h\rho h^\dagger}$ with smaller one-norm, and thus it holds that

$$\mathfrak{R}(h\rho h^\dagger) \leq \mathfrak{R}(\rho), \quad \forall \rho, \quad \forall h \in \text{Cl}_n. \quad (40)$$

(b) Pauli measurements. We consider the measurement of a Pauli observable T_a on a quantum state ρ . Denote by ρ_{a,s_a} the normalized postmeasurement states for the outcomes $s_a = 0, 1$, respectively. We have to show that, for all n , for all $a \in \mathbb{Z}_2^n \times \mathbb{Z}_2^n$ and all n -qubit states ρ it holds that

$$p_a(0) \mathfrak{R}(\rho_{a,0}) + p_a(1) \mathfrak{R}(\rho_{a,1}) \leq \mathfrak{R}(\rho). \quad (41)$$

With Eq. (33), we can write $p_a(0) W_{\rho_{a,0}} = W_+ + \overline{W}_+$, and $p_a(1) W_{\rho_{a,1}} = W_- + \overline{W}_-$, where

$$\begin{aligned} W_+(\Omega', \gamma') &:= \sum_{(\Omega, \gamma) \in \mathcal{V}} \frac{W_\rho(\Omega, \gamma)}{2} \delta_{a \in \Omega} \delta_{\gamma(a), 0} (\delta_{(\Omega', \gamma'), (\Omega, \gamma)} \\ &\quad + \delta_{(\Omega', \gamma'), (\Omega, \gamma + [a, \cdot])}), \\ W_-(\Omega', \gamma') &:= \sum_{(\Omega, \gamma) \in \mathcal{V}} \frac{W_\rho(\Omega, \gamma)}{2} \delta_{a \in \Omega} \delta_{\gamma(a), 1} (\delta_{(\Omega', \gamma'), (\Omega, \gamma)} \\ &\quad + \delta_{(\Omega', \gamma'), (\Omega, \gamma + [a, \cdot])}), \\ \overline{W}_+(\Omega', \gamma') &:= \sum_{(\Omega, \gamma) \in \mathcal{V}} \frac{W_\rho(\Omega, \gamma)}{2} \overline{\delta}_{a \in \Omega} \delta_{(\Omega', \gamma'), (\Omega \times a, \gamma \times (s_a = 0))}, \\ \overline{W}_-(\Omega', \gamma') &:= \sum_{(\Omega, \gamma) \in \mathcal{V}} \frac{W_\rho(\Omega, \gamma)}{2} \overline{\delta}_{a \in \Omega} \delta_{(\Omega', \gamma'), (\Omega \times a, \gamma \times (s_a = 1))}. \end{aligned} \quad (42)$$

From now on, denote by W_ρ the optimal representation for ρ w.r.t. one-norm, i.e., $\mathfrak{R}(\rho) = \|W_\rho\|_1$. With the triangle inequality, and the fact that the functions $W_{\rho_{a,s_a}}$ induced from the optimal W_ρ through Eq. (42) need not be optimal for the states ρ_{a,s_a} w.r.t. their one-norm, it holds that $p_{a,0} \mathfrak{R}(\rho_{a,0}) \leq \|W_+\|_1 + \|\overline{W}_+\|_1$, and $p_{a,1} \mathfrak{R}(\rho_{a,1}) \leq \|W_-\|_1 + \|\overline{W}_-\|_1$, hence

$$\begin{aligned} p_{a,0} \mathfrak{R}(\rho_{a,0}) + p_{a,1} \mathfrak{R}(\rho_{a,1}) \\ \leq \|W_+\|_1 + \|W_-\|_1 + \|\overline{W}_+\|_1 + \|\overline{W}_-\|_1. \end{aligned} \quad (43)$$

With Eq. (42) we find that

$$\begin{aligned} \|W_+\|_1 + \|W_-\|_1 \\ &= \sum_{(\Omega', \gamma') \in \mathcal{V}} \frac{\delta_{a \in \Omega'}}{2} |W_\rho(\Omega', \gamma') + W_\rho(\Omega', \gamma' + [a, \cdot])| \\ &\leq \sum_{(\Omega', \gamma') \in \mathcal{V}} \delta_{a \in \Omega'} |W_\rho(\Omega', \gamma')|, \end{aligned}$$

where in the second line we used the triangle inequality again. Furthermore, performing the summation over all $(\Omega', \gamma') \in \mathcal{V}$ first, we obtain

$$\|\overline{W}_+\|_1 = \|\overline{W}_-\|_1 \leq \sum_{(\Omega, \gamma) \in \mathcal{V}} \frac{\overline{\delta}_{a \in \Omega}}{2} |W_\rho(\Omega, \gamma)|.$$

Inserting the last two relations into inequality (43), we arrive at

$$p_{a,0} \mathfrak{R}(\rho_{a,0}) + p_{a,1} \mathfrak{R}(\rho_{a,1}) \leq \|W_\rho\|_1.$$

Since $\mathfrak{R}(\rho) = \|W_\rho\|_1$ by assumption, Eq. (41) follows. ■

TABLE IV. Robustness values of selected magic states. For robustness of magic (\mathfrak{R}_S), also see Ref. [45].

State	\mathfrak{R}	\mathfrak{R}_S
$ H\rangle^{\otimes 2}$	1.000	1.747
$ T\rangle^{\otimes 2}$	1.000	2.232
$ H\rangle^{\otimes 3}$	1.283	2.219
$ T\rangle^{\otimes 3}$	1.385	3.098
$ Hoggar\rangle$	1.800	3.800

D. Numerical results

In Table IV and Fig. 7 we present numerical values⁴ for the robustness of various magic states, and compare them to robustness of magic as defined by Howard and Campbell [27]. Table IV summarizes the robustness comparisons for the common magic states, as well as the maximal-robustness Hoggar state [27]. In Fig. 7 we plot the robustness against the stabilizer state robustness for three qubits, as a function of rotation angle. Note the wide and almost flat—though not perfectly flat—plateaus of robustness \mathfrak{R} in the vicinity of stabilizer states.

E. Curious resurgence of 4^n -dimensional phase space

Numerical calculations of robustness for various quantum states revealed an unexpected feature. Namely, the optimal quasiprobability distribution W_ρ w.r.t. Eq. (37) for a given n -qubit state ρ always was nonzero only on 4^n phase space points or fewer. 4^n is only a tiny fraction of the whole phase space \mathcal{V} and furthermore the naive expectation if one were completely oblivious of the differences between even and odd d . However, the support of the optimal $W(\rho)$ depends on the state ρ . We can now explain the initially puzzling upper bound on the size of the support, 4^n .

The robustness \mathfrak{R} of a state ρ defined in Eq. (37) is the solution to the convex optimization problem

$$\min_{\mathbf{q}} \{ \|\mathbf{q}\|_1 : M\mathbf{q} = \mathbf{b} \}, \quad (44)$$

where $M_{i,j} = \text{Tr}(A_{\alpha_j} P_i)$, $b_i = \text{Tr}(\rho P_i)$, $\{\alpha_j : 1 \leq j \leq |\mathcal{V}|\}$ is an enumeration of the phase points and P_i are the n -qubit Pauli operators. For each variable q_j in Eq. (44), define two new variables $q_j^+ := \max(0, q_j)$ and $q_j^- := \max(0, -q_j)$. Then the convex optimization problem of Eq. (44) is equivalent to the standard form linear program

$$\min_{\mathbf{q}} \left\{ \sum_j q_j^+ + q_j^- : \tilde{M}\tilde{\mathbf{q}} = \mathbf{b}, \tilde{\mathbf{q}} \geq 0 \right\}, \quad (45)$$

where $\tilde{M} = [M \quad -M]$ and $\tilde{\mathbf{q}} = [(\mathbf{q}^+)^T \quad (\mathbf{q}^-)^T]^T$. This doubles the number of variables but does not change the number of equality constraints. Since we know this problem is feasible (any physical state can be written as an affine combination of phase point operators) and bounded (no physical

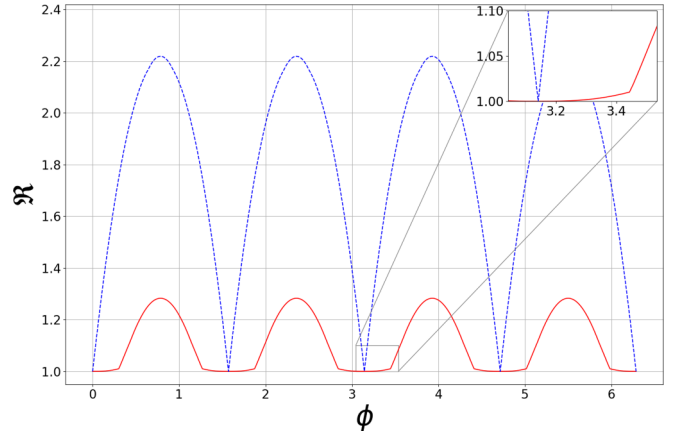


FIG. 7. Robustness \mathfrak{R} (solid line) and robustness of magic \mathfrak{R}_S (dotted line) for the state $|H(\phi)\rangle^{\otimes 3}$, cf. Eq. (14), as a function of ϕ . Highlighted is the region near a stabilizer state, at $\phi = \pi$.

state can have robustness less than 1), by the fundamental theorem of linear programming, for any physical state, Eq. (45) has a solution at a vertex of the feasible polytope [46].

Since Eq. (45) has an equality constraint for each n -qubit Pauli operator (including the identity), this means any state ρ has a robustness-minimizing expansion in phase point operators with no more than 4^n nonzero coefficients.

VIII. DISCUSSION

A. Stratonovich-Weyl correspondence

In the field of quantum optics, an important set of criteria for a proper quasiprobability distribution over a phase space is given by the Stratonovich-Weyl (SW) correspondence. Denote by $F_A^{(s)} : X \rightarrow \mathbb{R}$ the quasiprobability distribution corresponding to the operator A , with X the phase space and s a real parameter in the interval $[-1, 1]$. In the standard formalism for infinite-dimensional Hilbert spaces, $s = -1, 0, 1$ correspond to the Glauber-Sudarshan P , Wigner, and Husimi Q function, respectively. The SW correspondence is the following set of criteria on the $F_A^{(s)}$ [47]; also see Ref. [48],

(0) Linearity: $A \rightarrow F_A^{(s)}$ is a one-to-one linear map.

(1) Reality:

$$F_{A^\dagger}^{(s)}(u) = [F_A^{(s)}(u)]^*, \forall u \in X.$$

(2) Standardization:

$$\int_X d\mu(u) F_A^{(s)}(u) = \text{Tr} A.$$

(3) Covariance:

$$F_{g \cdot A}^{(s)}(u) = F_A^{(s)}(g^{-1}u), \quad g \in G,$$

with G the dynamical symmetry group.

(4) Traciality:

$$\int_X d\mu(u) F_A^{(s)}(u) F_B^{(-s)}(u) = \text{Tr} AB.$$

We now investigate to which extent these SW criteria apply to the present quasiprobability function W . There are two deviations. First, the present quasiprobability function W does

⁴Our calculations use the software packages CVXPY [53] and GUROBI [54].

not come with a parameter s ; there is only a single function W . This will affect the formulation of traciality. Second, the present mapping $A \rightarrow W_A$ is one-to-many, as we have noted in Sec. III A. The mapping is nonetheless linear, and $A + B$ can be represented as $W_A + W_B$.

The remaining SW conditions do apply. (1) Reality of W follows directly from the definition Eq. (4), since all A_Ω^γ are Hermitian. (2) Standardization: The definition Eq. (2) and property Eq. (3) of the phase point operators imply $\text{Tr} A_\Omega^\gamma = 1$, for all Ω and γ ; and standardization then follows from Eq. (4).

(3) Covariance holds for the entire Clifford group, as stated in Lemma 10. In fact, insisting on Clifford covariance leads to the nonuniqueness of W , namely, an overcomplete set of phase point operators is necessary to achieve Clifford covariance [49].

(4) Traciality. In the absence of a continuously varying parameter s , we define a dual quasiprobability function \tilde{W} in addition to W , to stand in for $F^{(-s)}$. For all Pauli operators T_a we have

$$\tilde{W}_{T_a}(\Omega, \gamma) := \begin{cases} (-1)^{\gamma(a)}, & a \in \Omega \\ 0, & a \notin \Omega \end{cases}$$

Since the n -qubit Pauli operators form an operator basis, \tilde{W} can be extended to all n qubit operators by linearity. With Eq. (34) we then have

$$\text{Tr} A\rho = \sum_{(\Omega, \gamma) \in \mathcal{V}} \tilde{W}_A(\Omega, \gamma) W_\rho(\Omega, \gamma).$$

We thus satisfy the SW criteria (1)–(4).

To conclude, we reiterate that for the present purpose of classically simulating QCM, a crucial property of W is positivity preservation under Pauli measurement. This property has no counterpart in the Stratonovich-Weyl correspondence.

B. Probabilistic hidden variable model

In the case of odd d [14], there is a third equivalent indicator of classicality, next to positivity of the initial Wigner function and the efficiency of classical simulation of QCM by sampling. Namely, a positive Wigner function is equivalent to a noncontextual hidden variable model (HVM) with deterministic value assignments [16]. This triple coincidence cannot be replicated in $d = 2$, because, for $n \geq 2$ all quantum states—even the completely mixed state—are contextual [16].

One interpretation of this situation is that contextuality, i.e., the unviability of noncontextual HVMs, is not sufficiently tight a criterion to reveal genuine quantumness. A more stringent marker is required, which (i) classifies the present HVM as classical, and (ii) for QCM in odd d reduces to contextuality. At present, we have no suggestion for this more restrictive notion of quantumness. However, we point to a hidden variable model that is illustrative of the shifted quantum-to-classical boundary in the multiqubit case, and we propose it for further study.

Namely, when positive, the quasiprobability distribution W can be considered an HVM. While classified as contextual by the common definitions, it shares many features with noncontextual HVMs.

This HVM consists of a triple $(\Lambda, \{h^\lambda\}, p_\lambda)$ where $\Lambda = \mathcal{V}$ or \mathcal{V}_M , h^λ is a compatible family of distributions on the set of

outcomes on contexts and p_λ is a probability distribution on the set Λ of hidden variables. For each $\alpha = (\Omega, \gamma)$ we define h^α by

$$h_I^\alpha(s) = \text{Tr}(P_s A_\Omega^\gamma). \quad (46)$$

Therein I is any isotropic subspace, $s : I \rightarrow \mathbb{Z}_2$ is a function, and P_s is the projector corresponding to the outcome. Note that $P_s = 0$ if $ds \neq \beta$.

It is useful to state the probability distributions $h_I^\alpha(\cdot)$ in their explicit form. Let P_s denote the projector corresponding to the noncontextual value assignment $s : I \rightarrow \mathbb{Z}_2$. Then we have

$$h_I^{(\Omega, \gamma)}(s) = \frac{|I \cap \Omega|}{|I|} \delta_{s|_{I \cap \Omega}, \gamma|_{I \cap \Omega}}. \quad (47)$$

From Eq. (47) we see that the value assignments in our HVM are generally probabilistic; only in the special case of $I \subset \Omega$ they become deterministic. Further, the $\{h_I^\alpha\}$ form compatible families,

$$h_I^\alpha|_{I \cap I'} = h_{I'}^\alpha|_{I \cap I'}, \quad \forall I, I', \quad \forall \alpha \in \mathcal{V}.$$

When applicable, this HVM reproduces the predictions of quantum mechanics (cf. Theorem 3) for measurements of Pauli observables, in single contexts or arbitrary measurement sequences.

We argue that the HVM of Eq. (46) is classical. It is an HVM with partial value assignments, with deterministic values for some observables and random values for others. The only resource this HVM uses beyond those required by noncontextual HVMs with deterministic value assignments is that of classical uniform randomness (in the evaluation of value assignments). Such use of randomness should not render the present HVM genuinely quantum.

And yet, the HVM of Eq. (46) is (a) contextual in the sense of Abramsky and Brandenburger [50], (b) preparation and transformation contextual, as well as measurement-noncontextual, in the sense of Spekkens [51], and (c) contextual for sequences of transformations in the recently introduced sense of Mansfield and Kashefi [52].

To summarize, we have described a hidden variable model corresponding to positive quasiprobabilities W . By this correspondence, the HVM is considered classical from the quantum optics perspective. It is also classical from the computational perspective, as it leads to efficient classical simulation of QCM (for applicable magic states). And yet this HVM is contextual, per the definitions commonly applied. As such, the present HVM may serve as a reference point for a refined foundational notion of quantumness that goes beyond contextuality.

IX. CONCLUSION

We have introduced a quasiprobability distribution W over generalized phase space, which is defined for any number n of qudits with any number d of levels. For multiqubit systems with odd local dimension d , W reduces to the familiar Wigner function for finite-dimensional systems defined by Gross [29]. For even d , the phase space is enlarged and W becomes nonunique. Importantly, also for $d = 2$ (the multiqubit case), W has the property that a positive quasiprobability function

remains positive under all Pauli measurements. This property is crucial for classical simulation algorithms of quantum computation with magic states (QCM) by sampling.

Once this fundamental property is established, it is natural to investigate the efficiency (or nonefficiency) of classical simulation in the various regimes and resource theories characterizing QCM. Here we have treated the canonical questions that arise in this context: we have devised an efficient classical simulation of QCM for $W \geq 0$ and clarified the relation to the qubit stabilizer formalism. Namely, the present method for efficient classical simulation of QCM strictly contains the stabilizer method. It applies to all mixtures of stabilizer states, but in addition to certain states outside the stabilizer polytope. We have further characterized the hardness of classical simulation for $W < 0$ in terms of a robustness measure, and established this robustness is a monotone under the free operations of QCM.

In summary, we arrive at a resource perspective of QCM on qubits that closely resembles the corresponding picture for odd dimension d . However, there are two deviations. First, the phase space on which the quasiprobability function W is defined has a far more intricate structure for $d = 2$ than for odd d . Second, for $d = 2$ the hidden variable model (HVM) induced by any non-negative quasiprobability function W_ρ is contextual, as a consequence of Mermin's square.

The latter observation leads to a puzzle. The HVM induced for positively representable states ρ is classified as "classical" from the perspectives of quantum optics ($W_\rho \geq 0$) and computer science (classical simulation is efficient), but it is classified as "quantum" from the perspective of contextuality.

In this regard, we have argued (also see Ref. [16]) that in multiqubit QCM, contextuality is not suitable as an indicator of genuine quantumness. We have proposed the notion of "HVM with partial noncontextual value assignments" in which classicality and contextuality coexist.

ACKNOWLEDGMENTS

We thank Piers Lillystone (J.B.-V., C.O., R.R., E.T.) and Shane Mansfield (J.B.-V.) for discussion. C.O., R.R., E.T., M.Z. are funded by NSERC, and R.R. acknowledges funding from Cifar. J.B.V. acknowledges funding from the ERC (TAQ 307498) project and the European Unions Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement No. 754446, and from the UGR Research and Knowledge Transfer Found Athena3i.

APPENDIX A: PROOF OF LEMMA 1

Recall that, unlike most material in this paper, Lemma 1 holds for all local dimensions d .

Proof of Lemma 1. Consider two sets, $\Omega, \tilde{\Omega} \in \mathcal{V}$, such that $\Omega \subset \tilde{\Omega}$, and the phase point operator A_Ω^γ according to Eq. (2). Furthermore, denote by $\tilde{\Gamma}$ the set of value assignments $\tilde{\gamma} : \tilde{\Omega} \rightarrow \mathbb{Z}_d$ that satisfy the constraint

$$\tilde{\gamma}|_\Omega = \gamma.$$

Then $\tilde{\Gamma}$ is the coset of a vector space U . This is the first fact we prove. Write $\tilde{\gamma} = \tilde{\gamma}_0 + \eta$, where $\tilde{\gamma}_0 \in \tilde{\Gamma}$ is some reference function, and the functions $\eta \in U$ all satisfy

$$d\eta = 0, \quad (\text{A1a})$$

$$\eta|_\Omega = 0. \quad (\text{A1b})$$

The condition of Eq. (A1a) need be satisfied only for commuting pairs of elements in $\tilde{\Omega}$. From Eq. (A1) it follows that if $\eta, \eta' \in U$ then $c\eta + c'\eta' \in U$, for all $c, c' \in \mathbb{Z}_d$. Hence U is indeed a vector space, as claimed.

Key is the relation

$$A_\Omega^\gamma = \frac{1}{|\tilde{\Gamma}|} \sum_{\tilde{\gamma} \in \tilde{\Gamma}} A_{\tilde{\Omega}}^{\tilde{\gamma}}, \quad (\text{A2})$$

which we now prove, armed with the previous observation. Using the definition of the phase point operators, we start expanding the r.h.s. of Eq. (A2):

$$\begin{aligned} \frac{1}{|\tilde{\Gamma}|} \sum_{\tilde{\gamma} \in \tilde{\Gamma}} A_{\tilde{\Omega}}^{\tilde{\gamma}} &= \frac{1}{|\tilde{\Gamma}|} \sum_{\tilde{\gamma} \in \tilde{\Gamma}} \frac{1}{d^n} \sum_{a \in \tilde{\Omega}} \omega^{\tilde{\gamma}(a)} T_a \\ &= \frac{1}{|\tilde{\Gamma}|} \frac{1}{d^n} \sum_{a \in \tilde{\Omega}} \omega^{\tilde{\gamma}_0(a)} T_a \sum_{\eta \in U} \omega^{\eta(a)}. \end{aligned}$$

Now we consider two cases. (i) $a \in \Omega$. Then, with property Eq. (A1b),

$$\sum_{\eta \in U} \omega^{\eta(a)} = |U|, \quad \forall a \in \Omega. \quad (\text{A3})$$

Furthermore, note $|\tilde{\Gamma}| = |U|$.

(ii) $a \in \tilde{\Omega} \setminus \Omega$. There is at least one $\eta \in U$ with $\eta(a) \neq 0$. Since U is a vector space, it follows by character orthogonality that

$$\sum_{\eta \in U} \omega^{\eta(a)} = 0, \quad \forall a \in \tilde{\Omega} \setminus \Omega. \quad (\text{A4})$$

Inserting Eqs. (A3) and (A4) in the above expansion, and furthermore using property Eq. (A1b), we find

$$\frac{1}{|\tilde{\Gamma}|} \sum_{\tilde{\gamma} \in \tilde{\Gamma}} A_{\tilde{\Omega}}^{\tilde{\gamma}} = \frac{1}{d^n} \sum_{a \in \Omega} \omega^{\gamma(a)} T_a = A_\Omega^\gamma.$$

This proves Eq. (A2). Now, wlog. we may choose $\tilde{\Omega}$ to be maximal. Since by definition any set Ω is contained in some maximal set $\tilde{\Omega}(\Omega)$, we may convert any positive state expansion over \mathcal{V} into a positive state expansion over \mathcal{V}_M ,

$$\rho = \sum_{\Omega, \gamma} c(\Omega, \gamma) A_\Omega^\gamma = \sum_{\Omega, \gamma} \frac{c(\Omega, \gamma)}{|\tilde{\Gamma}_{\tilde{\Omega}(\Omega)}|} \sum_{\tilde{\gamma} \in \tilde{\Gamma}_{\tilde{\Omega}(\Omega)}} A_{\tilde{\Omega}(\Omega)}^{\tilde{\gamma}}.$$

If the expansion coefficients on the l.h.s. are positive, so they are on the r.h.s. ■

APPENDIX B: PROOF OF LEMMA 7

Proof of Lemma 7. Statement (A): $(\Omega, \gamma + [a, \cdot]) \in \mathcal{V}, \forall a \in \Omega$. The set Ω does not change, and we need only to check the properties in Definition 4 that concern the function update, i.e., Eqs. (10) and (3).

Assume that $\gamma : \Omega \rightarrow \mathbb{Z}_2$ satisfies $d\gamma = \beta$ on Ω , i.e., $d\gamma(f) = \beta(f)$ for all faces $f \in F(\Omega)$. Consider any such face, with its boundary ∂f consisting of the edges c , d , and $c + d$. By definition of $F(\Omega)$ it holds that $c, d, c + d \in \Omega$. Then, with all addition mod 2,

$$\begin{aligned} d(\gamma + [a, \cdot])(f) &= d\gamma(f) + [a, \cdot](\partial f) \\ &= d\gamma(f) + [a, c] + [a, d] + [a, c + d] \\ &= d\gamma(f) \\ &= \beta(f). \end{aligned}$$

Thus, $\gamma + [a, \cdot]$ satisfies Eq. (10).

Furthermore, assume that γ satisfies Eq. (3). Then $(\gamma + [a, \cdot])(0) = \gamma(0) + [a, 0] = \gamma(0)$. Hence, $\gamma + [a, \cdot]$ satisfies Eq. (3).

Statement (B): $(\Omega \times a, \gamma \times s_a) \in \mathcal{V}$, $\forall a \notin \Omega$ and $s_a \in \mathbb{Z}_2$. There are four items to check in Definition 4, namely, (I) $\Omega \times a$ is closed under inference, (II) $\Omega \times a$ is noncontextual, (III) $\gamma \times s_a$ satisfies Eq. (10), and (IV) $\gamma \times s_a$ satisfies Eq. (3).

(I): Consider $c, d \in \Omega \times a$, with $[c, d] = 0$, and denote $c' = c + a$, $d' = d + a$. There are three subcases. (i) $c, d \in \Omega_a$. Then, $c + d \in \Omega_a$, since Ω_a is closed under inference by Lemma 6. Thus, $c + d \in \Omega \times a$.

(ii) $c \in \Omega_a$, $d \notin \Omega_a$. By construction of $\Omega \times a$, $d' \in \Omega_a$. Thus, $c + d = c + (d' + a) = (c + d') + a$. Now, since $[c, d] = 0$ by assumption and $[c, a] = 0$ ($c \in \Omega_a$) it follows that $[c, d'] = 0$. Since Ω_a is closed by Lemma 6, it holds that $c + d' \in \Omega_a$. By construction of $\Omega \times a$, $c + d = (c + d') + a \in \Omega \times a$.

(iii) $c, d \notin \Omega_a$. By construction of $\Omega \times a$, $c', d' \in \Omega_a$. Thus, $c + d = (c' + a) + (d' + a) = c' + d'$, and further $[c', d'] = 0$. Since Ω_a is closed under inference by Lemma 6, $c' + d' = c + d \in \Omega_a$. Thus, $c + d \in \Omega \times a$.

Thus in all three cases, $c, d \in \Omega \times a$, with $[c, d] = 0$, implies $c + d \in \Omega \times a$. Hence, $\Omega \times a$ is closed under inference.

(II): Assume that $d\gamma = \beta$ on Ω , and consider a triple of edges $c, d, c + d \in \Omega \times a$ with $[c, d] = 0$. Then either (i) all or (ii) one of these edges are in the component Ω_a .

(i) $c, d, c + d \in \Omega_a$. Since $\Omega_a \subset \Omega$ and with Eq. (25a), it holds that $d(\gamma \times s_a)(c, d) = d\gamma(c, d) = \beta(c, d)$.

(ii) W.l.o.g. assume that $c \in \Omega_a$ and $d, c + d \notin \Omega_a$, and denote $c' = c + a$, $d' = d + a$ as before. Then, for the face $f = (c, d)$ with boundary ∂f consisting of the edges c , d and $c + d$,

$$\begin{aligned} d(\gamma \times s_a)(f) &= \gamma \times s_a(c) + \gamma \times s_a(d) + \\ &\quad \gamma \times s_a(c + d) \\ &= \gamma(c) + \gamma(d') + \gamma(c + d') + \\ &\quad \beta(a, d) + \beta(a, c + d) \\ &= \beta(c, d') + \beta(a, d) + \beta(a, c + d) \\ &= \beta(c, d). \end{aligned}$$

Therein, in the second line we have used Eq. (25), in the third line Eq. (10), in the fourth line Eq. (28), and in the fourth line $d\beta(a, d, c) = 0$, cf. Eqs. (7) and (11).

(II): Per Definition (3), $\Omega \times a$ is noncontextual if there is a function $\tau : \Omega \times a \rightarrow \mathbb{Z}_2$ that satisfies $d\tau = \beta$. We

have explicitly constructed such a function in (III) above, $\tau := \gamma \times s_a$.

(IV): Assume that $\gamma : \Omega \rightarrow \mathbb{Z}_2$ satisfies Eq. (3). Since $0 \in \Omega_a$ for all cnc sets Ω , with Eq. (25a) it follows that $\gamma \times s_a(0) = \gamma(0)$, and hence $\gamma \times s_a$ also satisfies Eq. (3). ■

APPENDIX C: PROOF OF LEMMA 9

Proof of Lemma 9. Recall from Lemma 3 that each set Ω can be written in the form $\Omega = \bigcup_{k=1}^{\xi(\Omega)} I_k$, where each I_k is an isotropic subspace, $I_k = \langle a_k, \tilde{I} \rangle$, $a_k \in E$. Therefore, for all $(\Omega, \gamma) \in \mathcal{V}$, it holds that

$$A_\Omega^\gamma = \left(\sum_{k=1}^{\xi(\Omega)} A_{I_k}^{\gamma|_{I_k}} \right) - (\xi(\Omega) - 1) A_{\tilde{I}}^{\gamma|_{\tilde{I}}}.$$

Therein, the phase point operators appearing on the r.h.s. are all of the type $m = 0$, i.e., they correspond to stabilizer states. The Wigner function $\delta_{(\Omega, \gamma)}$ representing the operator A_Ω^γ can thus be expanded as

$$\delta_{(\Omega, \gamma)} = \left(\sum_{k=1}^{\xi(\Omega)} \delta_{(I_k, \gamma|_{I_k})} \right) - (\xi(\Omega) - 1) \delta_{(\tilde{I}, \gamma|_{\tilde{I}})}.$$

Denote by $\|\cdot\|_1$ the one-norm of the expansion in terms of phase point operators A_Ω^γ , and by $\|\cdot\|_{1,S}$ the one-norm of the expansion in terms of (density matrices of) stabilizer states. With the last equation, the triangle inequality, $\|\delta_{(I_k, \gamma|_{I_k})}\|_{1,S} = \|\delta_{(\tilde{I}, \gamma|_{\tilde{I}})}\|_{1,S} = 1$, and $\xi(\Omega) \leq 2n + 1$ for all cnc sets Ω (cf. Lemma 3), it follows that

$$\|\delta_{(\Omega, \gamma)}\|_{1,S} \leq 4n + 1. \quad (\text{C1})$$

Now, for any given state ρ consider the optimal representation W_ρ , i.e., the one with minimal norm $\|W_\rho\|_1$. Then

$$\begin{aligned} \mathfrak{R}_S(\rho) &\leq \|W_\rho\|_{1,S} = \left\| \sum_{(\Omega, \gamma) \in \mathcal{V}} W_\rho(\Omega, \gamma) \delta_{(\Omega, \gamma)} \right\|_{1,S} \\ &\leq (4n + 1) \sum_{(\Omega, \gamma) \in \mathcal{V}} |W_\rho(\Omega, \gamma)| \\ &= (4n + 1) \|W_\rho\|_1 \\ &= (4n + 1) \mathfrak{R}(\rho). \end{aligned}$$

Therein, in the first line we have an inequality because the representation W_ρ of ρ is optimized for $\|W_\rho\|_1$, not necessarily for $\|W_\rho\|_{1,S}$. The third line follows by the triangle inequality and Eq. (C1), and the fifth line holds as an equality because W_ρ , per assumption, was chosen to minimize $\|W_\rho\|_1$.

This proves the right half of Eq. (38). The left half, $\mathfrak{R}(\rho) \leq \mathfrak{R}_S(\rho)$, follows from the fact that all stabilizer states correspond to phase point operators of type $m = 0$. Hence, an expansion in terms of stabilizer states induces an expansion in terms of phase point operators A_Ω^γ , with the same nonzero coefficients. ■

APPENDIX D: COMPUTING W REPRESENTATIONS OF MANY COPIES OF MAGIC STATES

Here we describe how to construct valid quasiprobabilities $W_{\mu^{\otimes n}}$ for n copies of a magic state μ , at bounded compu-

tational cost. As with robustness of magic [27], we merge expansions for small numbers of magic states into valid expansions for larger numbers of copies.

Denote by Ω_n^m cnc sets Ω with parameters $n, m \leq n$, and choose the phases ϕ in Eq. (1) such that

$$T_{a+b} = T_a \otimes T_b, \quad \forall a \in \Omega_{n_1}^{m_1}, b \in \Omega_{n_2}^{m_2}. \quad (\text{D1})$$

Here we identified a and b as elements of $\mathbb{Z}_d^{2(n_1+n_2)}$ by writing $((a_X, 0), (a_Z, 0))$ and $((0, b_X), (0, b_Z))$, respectively. We then have the following result.

Lemma 11. Be $\Omega_{n_1}^{m_1}$ and $\Omega_{n_2}^0$ two cnc sets with parameters $n_1, m_1 \leq n$, and $n_2, m_2 = 0$, respectively. Then $A_{\Omega_{n_1}^{m_1} \oplus \Omega_{n_2}^0}^\gamma := A_{\Omega_{n_1}^{m_1}}^{\gamma_1} \otimes A_{\Omega_{n_2}^0}^{\gamma_2}$, with the function $\gamma : \Omega_{n_1}^{m_1} \oplus \Omega_{n_2}^0 \rightarrow \mathbb{Z}_2$ defined by

$$\gamma(a_1 + a_2) := \gamma_1(a_1) + \gamma_2(a_2), \quad (\text{D2})$$

for all $a_1 \in \Omega_{n_1}^{m_1}, a_2 \in \Omega_{n_2}^0$ is a valid phase point operator on $n_1 + n_2$ qubits.

Proof of Lemma 11. We need to verify the properties of Definition 4, namely, (a) that $\Omega_{n_1}^{m_1} \oplus \Omega_{n_2}^0$ is cnc, and (b) that the function γ defined in Lemma 11 satisfies Eq. (10), i.e., $d\gamma = \beta$.

Regarding (a), with Eq. (15), $\Omega_{n_1}^{m_1} = \bigcup_{k=1}^{m_1} \langle a_k, \tilde{I} \rangle$, and $\Omega_{n_2}^0 = I$, with \tilde{I}, I isotropic subspaces. Then

$$\Omega_{n_1}^{m_1} \oplus \Omega_{n_2}^0 = \bigcup_{k=1}^{m_1} \langle a_k, \tilde{I} \oplus I \rangle,$$

and $\tilde{I} \oplus I$ is also an isotropic subspace. Hence the set $\Omega_{n_1}^{m_1} \oplus \Omega_{n_2}^0$ is cnc by Lemma 3.

Regarding (b), we need to show that for all $a, b \in \Omega_{n_1}^{m_1} \oplus \Omega_{n_2}^0$ with $[a, b] = 0$ it holds that $\gamma(a + b) + \gamma(a) + \gamma(b) = \beta(a, b)$. To this end, for any given commuting pair a, b , we split

$$a = a_1 + a_2, \quad b = b_1 + b_2,$$

with $a_1, b_1 \in \Omega_{n_1}^{m_1}, a_2, b_2 \in \Omega_{n_2}^0$. The decompositions are unique. Since a and b commute, $[a_1, b_1] + [a_2, b_2] = 0$. Furthermore, a_2 commutes with b_2 , since $\Omega_{n_2}^0$ is an isotropic subspace. Thus,

$$[a_1, b_1] = [a_2, b_2] = 0. \quad (\text{D3})$$

Further, Eq. (D1) implies that

$$\beta(c, d) = 0, \quad \forall c \in \Omega_{n_1}^{m_1}, d \in \Omega_{n_2}^0. \quad (\text{D4})$$

Now we rewrite

$$\begin{aligned} \gamma(a + b) &= \gamma[(a_1 + b_1) + (a_2 + b_2)] \\ &= \gamma_1(a_1 + b_1) + \gamma_2(a_2 + b_2) \\ &= \gamma_1(a_1) + \gamma_1(b_1) + \gamma_2(a_2) + \gamma_2(b_2) \\ &\quad + \beta(a_1, b_1) + \beta(a_2, b_2) \\ &= \gamma(a_1 + a_2) + \gamma(b_1 + b_2) + \beta(a_1, b_1) \\ &\quad + \beta(a_2, b_2) + \beta(a_1, a_2) + \beta(b_1, b_2) \\ &= \gamma(a) + \gamma(b) + \beta(a_1 + a_2, b_1 + b_2) \\ &\quad + d\beta(a_1, a_2, b_1 + b_2) + d\beta(a_2, b_2, b_1) \\ &\quad + d\beta(a_1, b_1, a_2 + b_2) \\ &= \gamma(a) + \gamma(b) + \beta(a, b). \end{aligned}$$

Therein, in the second line we have used the definition of γ in Lemma 11. In the third line Eq. (D3), and in the fourth line the definition of γ again. In the fifth line we have used Eq. (11), Eq. (8), and Eq. (D4) on $\beta(a_1 + b_1, a_2 + b_2)$. In the last line we used Eq. (7) ($d\beta = 0$). ■

Denote by $W_\sigma^{(0)}$ an expansion Eq. (4) of the state σ , but containing only phase point operators with parameter $m = 0$, i.e., an expansion into stabilizer states. W_ρ is a valid expansion of ρ , according to Eq. (4). Then it follows from Lemma 11 that

$$W_{\rho \otimes \sigma} := W_\rho \otimes W_\sigma^{(0)} \quad (\text{D5})$$

is a valid expansion of $\rho \otimes \sigma$.

Let k be the largest integer for which decompositions $W_{\mu^{\otimes k}}$ and $W_{\mu^{\otimes k}}^{(0)}$ are obtainable. Then, with Eq. (D5), $W_{\mu^{\otimes n}} = W_{\mu^{\otimes k}} \otimes W_{\mu^{\otimes n-k}}^{(0)}$, and the second factor may be further decomposed as $W_{\mu^{\otimes n-k}}^{(0)} = W_{\mu^{\otimes k}}^{\otimes(n/k-1)}$, if k divides n . Thus, we arrive at an explicit decomposition for $\mu^{\otimes n}$, with one-norm

$$\|W_{\mu^{\otimes n}}\| = \mathfrak{R}(\mu^{\otimes k}) \mathfrak{R}_S(\mu^{\otimes k})^{n/k-1}.$$

Thus the reduction to blocks of k magic states familiar from the stabilizer case [27] can be applied in the present setting as well. By Lemma 9, the resulting one-norm is lower by a constant factor than of the corresponding expansion into stabilizer states.

-
- [1] P. Ehrenfest, Bemerkung über die angenäherte Gültigkeit der klassischen Mechanik innerhalb der Quantenmechanik, *Z. Physik* **45**, 455 (1927).
 [2] A. Einstein, B. Podolsky, and N. Rosen, Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.* **47**, 777 (1935).
 [3] E. Schrödinger, Die gegenwärtige Situation in der Quantenmechanik, *Naturwissenschaften* **23**, 807 (1935).

- [4] R. Feynman, Simulating physics with computers, *Intl. J. Theor. Phys.* **21**, 467 (1982).
 [5] E. Bernstein and U. Vazirani, Quantum complexity theory, *SIAM J. Comput.* **26**, 1411 (1997). First appeared in ACM STOC (1993).
 [6] D. Deutsch, Quantum theory, the Church-Turing principle and the universal quantum computer, *Proc. Roy. Soc. London A* **400**, 97 (1985).

- [7] S. Bravyi and A. Kitaev, Universal quantum computation with ideal Clifford gates and noisy ancillas, *Phys. Rev. A* **71**, 022316 (2005).
- [8] E. Wigner, On the quantum correction for thermodynamic equilibrium, *Phys. Rev.* **40**, 749 (1932).
- [9] R. L. Hudson, When is the Wigner quasi-probability density non-negative? *Rep. Math. Phys.* **6**, 249 (1974).
- [10] V. Buzek, A. Vidiella-Barranco, and P. L. Knight, Superpositions of coherent states: Squeezing and dissipation, *Phys. Rev. A* **45**, 6570 (1992).
- [11] J. Bell, On the Einstein Podolsky Rosen paradox, *Physics* **1**, 195 (1964).
- [12] S. Kochen and E. P. Specker, The problem of hidden variables in quantum mechanics, *J. Math. Mech.* **17**, 59 (1967).
- [13] N. D. Mermin, Hidden variables and the two theorems of John Bell, *Rev. Mod. Phys.* **65**, 803 (1993).
- [14] V. Veitch, C. Ferrie, D. Gross, and J. Emerson, Negative quasi-probability as a resource for quantum computation, *New J. Phys.* **14**, 113011 (2012).
- [15] A. Mari and J. Eisert, Positive Wigner Functions Render Classical Simulation of Quantum Computation Efficient, *Phys. Rev. Lett.* **109**, 230503 (2012).
- [16] M. Howard, J. J. Wallman, V. Veitch, and J. Emerson, Contextuality supplies the magic for quantum computation, *Nature (London)* **510**, 351 (2014).
- [17] D. Gottesman, Stabilizer codes and quantum error correction, Ph.D. thesis, California Institute of Technology, 1997.
- [18] N. Delfosse, C. Okay, J. Bermejo-Vega, D. E. Browne, and R. Raussendorf, Equivalence between contextuality and negativity of the Wigner function for qudits, *New J. Phys.* **19**, 123024 (2017).
- [19] 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).
- [20] R. Raussendorf, D. E. Browne, N. Delfosse, C. Okay, and J. Bermejo-Vega, Contextuality as a resource for qubit quantum computation, *Phys. Rev. A* **95**, 052334 (2017).
- [21] J. Bermejo-Vega, N. Delfosse, D. E. Browne, C. Okay, and R. Raussendorf, Contextuality as a Resource for Models of Quantum Computation with Qubits, *Phys. Rev. Lett.* **119**, 120505 (2017).
- [22] L. Kocia and P. Love, Discrete Wigner formalism for qubits and non-contextuality of Clifford gates on qubit stabilizer states, *Phys. Rev. A* **96**, 062134 (2017).
- [23] E. F. Galvao, Discrete Wigner functions and quantum computational speedup, *Phys. Rev. A* **71**, 042302 (2005).
- [24] M. Van den Nest, Classical simulation of quantum computation, the Gottesman-Knill theorem, and slightly beyond, *Quant. Inf. Comp.* **10**, 258 (2010).
- [25] M. Van den Nest, Simulating quantum computers with probabilistic methods, *Quant. Inf. Comp.* **11**, 784 (2011).
- [26] B. M. Terhal and D. P. DiVincenzo, Adaptive quantum computation, constant depth quantum circuits and Arthur-Merlin games, *Quant. Inf. Comp.* **4**, 134 (2004).
- [27] 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).
- [28] H. Pashayan, J. J. Wallman, and S. D. Bartlett, Estimating Outcome Probabilities of Quantum Circuits Using Quasiprobabilities, *Phys. Rev. Lett.* **115**, 070501 (2015).
- [29] D. Gross, Computational power of quantum many-body states and some results on discrete phase spaces, Ph.D thesis, Imperial College London, 2005.
- [30] D. Gross, Hudson's theorem for finite-dimensional quantum systems, *J. Math. Phys.* **47**, 122107 (2006).
- [31] K. S. Gibbons, M. J. Hoffman, and W. K. Wootters, Discrete phase space based on finite fields, *Phys. Rev. A* **70**, 062101 (2004).
- [32] A. Karanjai, J. J. Wallman, and S. D. Bartlett, Contextuality bounds the efficiency of classical simulation of quantum processes, [arXiv:1802.07744](https://arxiv.org/abs/1802.07744).
- [33] J. J. Wallman and S. D. Bartlett, Nonnegative subtheories and quasiprobability representations of qubits, *Phys. Rev. A* **85**, 062121 (2012).
- [34] A. Cabello, M. Gu, O. Gühne, and Z.-P. Xu, Optimal Classical Simulation of State-Independent Quantum Contextuality, *Phys. Rev. Lett.* **120**, 130401 (2018).
- [35] C. Okay, S. Roberts, S. D. Bartlett, and R. Raussendorf, Topological proofs of contextuality in quantum mechanics, *Quant. Inf. Comp.* **17**, 1135 (2017).
- [36] R. Raussendorf, Cohomological framework for contextual quantum computations, *Quant. Inf. Comp.* **19**, 1141 (2019).
- [37] M. Howard, E. Brennan, and J. Vala, Quantum contextuality with stabilizer states, *Entropy* **15**, 2340 (2013).
- [38] P. Rall, D. Liang, J. Cook, and W. Kretschmer, Simulation of qubit quantum circuits via Pauli propagation, *Phys. Rev. A* **99**, 062337 (2019).
- [39] W. M. Kirby and P. J. Love, Contextuality Test of the Nonclassicality of Variational Quantum Eigensolvers, *Phys. Rev. Lett.* **123**, 200501 (2019).
- [40] M. C. Golumbic, Trivially perfect graphs, *Discrete Math.* **24**, 105 (1978).
- [41] M. Aschbacher, *Finite Group Theory*, 2nd ed., Cambridge Studies in Advanced Mathematics Vol. 10 (Cambridge University Press, Cambridge, 2000).
- [42] P. Lillystone and J. Emerson, A contextual ψ -epistemic model of the n -qubit stabilizer formalism, [arXiv:1904.04268](https://arxiv.org/abs/1904.04268).
- [43] D. Gottesman, The Heisenberg representation of quantum computers, in *Proc. XXII Intl. Colloq. on Group Theoretical Methods in Physics*, edited by S. P. Corney, R. Delbourgo, and P. D. Jarvis (International Press, Cambridge, MA, 1999), pp. 32–43.
- [44] F. G. S. L. Brandao, and G. Gour, Reversible Framework for Quantum Resource Theories, *Phys. Rev. Lett.* **115**, 070503 (2015).
- [45] M. Heinrich and D. Gross, Robustness of magic and symmetries of the stabiliser polytope, *Quantum* **3**, 132 (2019).
- [46] J. Nocedal and S. Wright, *Numerical Optimization*, 2nd ed. (Springer Science and Business Media, New York, 2000).
- [47] R. L. Stratonovich, On distributions in representation space, *Zh. Eksp. Teor. Fiz.* **31**, 1012 (1956) [*Sov. Phys. JETP* **4**, 891 (1957)].
- [48] C. Brif and A. Mann, A general theory of phase-space quasiprobability distributions, *J. Phys. A* **31**, L9 (1998).

- [49] H. Zhu, Permutation Symmetry Determines the Discrete Wigner Function, *Phys. Rev. Lett.* **116**, 040501 (2016).
- [50] S. Abramsky and A. Brandenburger, The sheaf-theoretic structure of non-locality and contextuality, *New J. Phys.* **13**, 113036 (2011).
- [51] R. W. Spekkens, Contextuality for preparations, transformations, and unsharp measurements, *Phys. Rev. A* **71**, 052108 (2005).
- [52] S. Mansfield and E. Kashefi, Quantum Advantage from Sequential-Transformation Contextuality, *Phys. Rev. Lett.* **121**, 230401 (2018).
- [53] S. Diamond and S. Boyd, CVXPY: A Python-embedded modeling language for convex optimization, *J. Mach. Learn. Res.* **17**, 83 (2016).
- [54] Gurobi Optimization, LLC, Gurobi Optimizer Reference Manual, <http://www.gurobi.com>, 2018.