Phase space simulation method for quantum computation with magic states on qubits

Robert Raussendorf$^{1,2}$, Juani Bermejo-Vega$^3$, Emily Tyhurst$^4$, Cihan Okay$^{1,2}$, and Michael Zurel$^1$

1: Department of Physics & Astronomy, University of British Columbia, Vancouver, BC V6T1Z1, Canada
2: Stewart Blusson Quantum Matter Institute, University of British Columbia, Vancouver, BC, Canada
3: Freie Universität Berlin, Berlin, Germany
4: Department of Physics, University of Toronto, Toronto, ON, Canada

We propose a method for classical simulation of finite-dimensional quantum systems, based on sampling from a quasi-probability 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 quasi-probability 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 quasi-probability distributions, where it provides amplitude 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.

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-Podolosky-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 non-contextual 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 coincide [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. Non-contextual hidden variable models for QCM are unviable 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, negativ-
ity 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]. A positive Wigner function also implies a non-contextual hidden variable model; and, for \( n \geq 2 \) quantum systems, the reverse direction holds as well [16], [18]. The onset of negativity in the initial Wigner function, the appearance of contextuality, and the breakdown of efficient classical simulation by sampling all occur simultaneously. Wigner function negativity and contextuality of the magic states are necessary quantum computational resources.

As we noted, this picture only applies if the local Hilbert space dimension is odd. This excludes the full multi-qubit case, which arguably is the most important. Approaches to the full qubit case have been made, e.g. through the rebit scenario [19] and multi-qubit settings with operational restrictions [20], or by invoking a Wigner function over Grassmann variables [22], or multiple Wigner functions at once [23]. Common to these approaches is that they fail to efficiently simulate the evolution under general Clifford gates and Pauli measurements (subsequently “Clifford circuits”), unlike in the case of odd \( d \) [13]. This is a critical shortcoming, for two reasons. First, it means that these methods cannot efficiently simulate \( n \)-qubit QCM even in instances where magic states are absent. Second, Clifford circuits are efficiently simulable, for all \( d \), by another method—the stabilizer formalism [17]. The above approaches [19]–[23] do not subsume the stabilizer formalism, whereas the Wigner function method for odd \( d \) does.

An alternative approach is to define a quasiprobability function over stabilizer states [7], [24], [25], bypassing Wigner functions (also see [26]–[28] for probabilistic simulation). It has the advantage of efficiently simulating all Clifford circuits on positively represented states. For multi-qubit systems \((d = 2)\), it is so far unknown how the stabilizer method relates to Wigner functions, due to the above-described deficiencies of the latter.

In this paper, we provide the thus far missing phase space picture for QCM on multi-qubit systems. Central to our discussion is a new 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 adaption [29], [30] (also see [31]) of the original Wigner function [8], but for even \( d \), in particular \( d = 2 \), it requires a phase space of increased size. This new quasiprobability function has the property that, also in \( d = 2 \), all QCMs on positively representable states can be efficiently classically simulated. Further, this setting 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 \( W < 0 \), 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 non-contextual 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, hidden variable model and efficient classical simulation existing in odd \( d \). Central to our discussion is a novel 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, \( W_\rho \) is non-unique for any given quantum state \( \rho \). The set of phase point operators corresponding to \( W \) is over-complete.

(iii) If \( d \) is odd, 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 non-mixtures of stabilizer states which are positively represented by \( W \). Furthermore, for any quantum state \( \rho \), the l-norm of the optimal \( W_\rho \) is smaller or equal than the robust-
ness of magic $R_{\rho}(\rho)$. (Both robustness measures are instances of sum negativity [25].)

The following properties of $W$ for special values of $n$ (and $d = 2$) are also worth noting. (a) The Eight-state model [32] is a special case of $W$, namely for $n = 1$. (b) For Mermin’s square [13], the present simulation algorithm saturates the lower bound [33] 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 Section V). (II) If a quantum state $\rho$ has a non-negative function $W_\rho$, and $W_\rho$ can be efficiently sampled, then, for every Clifford circuit applied to $\rho$, the corresponding measurement statistics can be efficiently sampled (Theorem 3 in Section VII). In this sense, $W \geq 0$ leads to efficient classical simulation of the corresponding quantum computation. (III) For $d = 2$, 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 multi-qubit phase space are classified (Theorem 1 in Section IV). (IV) There exists a robustness measure $R$ which bounds the hardness of classical simulation of quantum computation with magic states, when $W_{\rho_{\text{init}}} < 0$ for the initial state $\rho_{\text{init}}$. $R$ is less than or equal to the robustness of magic (Lemma 10), and a monotone under Clifford unitaries and Pauli measurements (Theorem 4 in Section VIII).

B. Outline

The remainder of this paper is organized as follows. In Section 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 Section 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 Sections V and VI we turn to dynamics. In Section V we discuss the update of $W$ under Pauli measurement, and in Section VI the efficient classical simulation of QCM for positive $W$. Section VII describes the relation between non-negative quasiprobability distributions $W$ and hidden variable models. In Section VIII we turn to the case of $W_\rho < 0$. We discuss hardness of classical simulation, as well as the elements of a resource theory based on $W$. We conclude in Section IX.

III. THE QUASIPROBABILITY FUNCTION

A. Generalized phase space

In this section we introduce the generalized $n$-qudit phase space $V$, for any local Hilbert space dimension $d$, and a quasi-probability distribution $W : V \rightarrow \mathbb{R}$ living on it.

We choose a phase convention for the Pauli operators,

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

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 $\omega^k$, $k \in \mathbb{N}$, with $\omega := \exp(2\pi i/d)$.

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

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

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 $\Omega$ are an additional varying parameter, and the phase space thereby becomes larger.

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

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

The reason for imposing Eq. (3) is that it implies $\text{Tr}A_{\Omega,\gamma}^2 = 1$, for all $(\Omega, \gamma) \in V$. 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 the quasiprobability distribution $W_\rho$ is non-unique because the set of phase point operators of Eq. (4) is overcomplete.
The motivation for this definition is that if there exists a value assignment \( \gamma \). A set

\[ \Omega \]

is closed under inference by definition. And second, it is known that in odd dimension Pauli observables do not admit a parity-based contextuality proof. Thus, for the generalized phase space \( \mathcal{V} \), the only sets \( \Omega \) that matter are simultaneously closed and non-contextual. For short, we call such sets “cnc”.

### B. Maximal sets \( \Omega \)

The cnc sets \( \Omega \) 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 \( \Omega \) 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 \( \Omega \subseteq \tilde{\Omega} \).

We denote by \( \mathcal{V}_M \) the subset of \( \mathcal{V} \) constructed only from the maximal cnc sets \( \Omega \). Then, any quantum state \( \rho \) 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 \( \rho \) is positively representable w.r.t. \( \mathcal{V}_M \).

### Lemma 1

For any \( n \) and \( d \), a quantum state \( \rho \) 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 Section III C right below, and in the classification of cnc sets \( \Omega \) for the multi-qubit case in Section IV A.

### C. 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 the generalized phase space \( \mathcal{V} \) reduces to the standard phase space \( \mathcal{V} = \mathbb{Z}_d^{2n} \), which is the group of characters of \( E \). The quasiprobability function \( W \) becomes the standard Wigner function \( \mathcal{W} \) for odd finite-dimensional systems. Hence, the present quasiprobability function \( W \) is a generalization of the finite odd dimensional Wigner function \( \mathcal{W} \), which in turn is a descendant of the original Wigner function \( \mathcal{W} \).

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 non-contextual deterministic value assignments. These yield the functions \( \gamma \), satisfying the condition Eq. (5). \( E \) is thus non-contextual.
$E$ 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
\[
A^\gamma_E = \frac{1}{d^n} \sum_{a \in E} \omega^\gamma(a) T_a,
\]
with the functions $\gamma$ satisfying Eqs. (3) and (8). The former condition ensures that the identity operator appears with weight $1/d^n$ in the expansion (real and positive). The latter condition has $d^{2n}$ solutions for the functions $\gamma$ if $d$ is odd. For a suitable choice of $\phi$ in Eq. (1), it holds that $\beta = 0$ (odd $d$ only). The solutions for $\gamma$ then form a vector space
\[
V = \mathbb{Z}_d^{2n} \quad \text{for odd } d.
\]

D. 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—two strikingly simple contextuality proofs. Those settings embed most efficiently in rebits rather than qubits, which warrants their discussion 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. What makes these two treatments of the rebit case so different is that, 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.

We start the exploration of the $d = 2$ case with two examples that illustrate the concept of generalized phases space $V$. 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 [32], which consists of two standard 1-qubit Wigner functions tagged together. The Eight-state-model is an instance of the state expansion Eq. (4), and it contains only one set $\Omega$, namely
\[
\Omega_0 = \{0, x, y, z\},
\]
with $T_0 = I$, $T_x = X$, $T_y = Y$ and $T_z = Z$. It is easily checked that $\Omega_0$ is non-contextual and closed under inference (no inference possible). The value assignments $\gamma$ are constrained by Eq. (3), hence $\gamma(0) = 0$, and no constraints arise from Eq. (6) due to the lack of non-trivial commuting elements in $\Omega_0$. 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 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. Fig. 1 shows three distinct types of cnc sets $\Omega$. Type (a) is the union of two non-trivially intersecting isotropic subspaces (9 sets), type (b) is isotropic subspaces (6 sets), and type (c) is triples of anti-commuting elements, i.e., one from each row and column of the square (6 sets).

We make the following numerical observations: (i) Random sampling suggests that all 2-rebit states and most 2-qubit states are positively representable; see Table I. (ii) Two copies of the magic state measurements [20].

\[
|H(\phi)\rangle := (|0\rangle + e^{-i\phi}|1\rangle)/\sqrt{2}
\]
(can be positively represented, for all angles $\phi$. (iii) In Fig. 2 the region of positively representable density matrices of the form
\[
\rho(x, y) = \frac{1}{4}J_{12} + x(X_1X_2 + Z_1Z_2) - Y_1Y_2) + y(Z_1 + Z_2),
\]
for $x, y \in \mathbb{R}$, is displayed for three different methods; namely the stabilizer method, the hyper-octahedral
E. The cohomological viewpoint

The above Definitions 3 and 4 have a cohomological underpinning, which arises as follows: The partial value assignments \( \gamma \) and the function \( \beta \) are cochains in a chain complex, with Eqns. (6) and (8) constraining them. Eq. (6) says that \( \beta \) 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 \( \beta \) is non-trivial whereas for odd \( d \) it is trivial.

Eqns. (6) and (8) 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 11). These are central properties for the phase-space description of quantum computation with magic states, and they are all matters of cohomology. In addition, we note that the discussion of contextuality in measurement-based quantum computation (MBQC) employs the same cohomological structures; see [63].

The cohomological formulation is based on a chain complex \( C_n \) constructed from the \( n \)-qubit Pauli operators \( T_a \). The operator labels \( a \) define the edges of this complex; the faces of \( C_n \) correspond to commuting pairs \((a, b)\) and volumes \((a, b, c)\) to commuting triples. For details, the interested reader is referred to [35]. Here, we only state two basic topological properties of the present scenario.

Cohomology arises through the function \( \beta \) defined in Eq. (5), which is a 2-cochain on \( C_n \). As already noted, \( \beta \) is a 2-cocycle, with the cocycle condition \( d\beta = 0 \) enforced by Eq. (6). For any given volume \( v = (a, b, c) \), the coboundary \( d\beta \) evaluates on \( v \) to

\[
d\beta(a, b, c) := \beta(a, b) + \beta(a + b, c) - \beta(b, c) - \beta(a, b + c).
\]

(11)

Thus, Eq. (6) says that \( d\beta(v) = 0 \), for all volumes \( v \).

Eq. (8) in Definition 3 also has a cohomological interpretation, namely \( d\gamma = \beta|_{\Omega \times \Omega} \), with

\[
d\gamma(a, b) = \gamma(a) + \gamma(b) - \gamma(a + b),
\]

(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 Eqns. (11) and (12), as a short-hand to express Eqns. (9) 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. STRUCTURE OF THE PHASE SPACE $\mathcal{V}$

In this section, we look at the structure of the phase space $\mathcal{V}$ more closely. Namely, in Section IV A we classify the cnc sets $\Omega$, and for every $\Omega$ describe the sets $\Gamma(\Omega)$ of value assignments $\gamma$. In Section IV B we clarify the relation to the stabilizer formalism.

A. Classification of phase space points

Denote by $\Gamma(\Omega)$ the set of functions $\gamma : \Omega \rightarrow \mathbb{Z}_2$ that satisfy the constraint Eq. (8). Then, the following statement holds.

Lemma 2 For all sets $\Omega$ 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 $\eta + \eta' \in U(\Omega)$, 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 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 $\Omega$.

At this point, one question about the structure of $\mathcal{V}$ remains: Can the cnc sets $\Omega$ 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 $k = 1, \ldots, \xi$. For any number $n$ of qubits, the sets

$$\Omega = \bigcup_{k=1}^{\xi} I_k$$

are non-contextual and closed under inference.

Proof of Lemma 3. Existence. The sets $\Omega$ of Eq. (13) 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 $m$,

$$C_{2j-1} = I_{1,j-1}X_jZ_{j+1}Z_{j+2}Z_{m-1}Z_m,$$

$$C_{2j} = I_{1,j-1}Y_jZ_{j+1}Z_{j+2}Z_{m-1}Z_m,$$

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

$$C_{2m+1} = Z_1Z_2\ldots Z_{m-1}Z_m.$$

Further, be $\tilde{I}$ the isotropic subspace corresponding to a stabilizer state supported on the $n - m$ qubits numbered $m + 1, \ldots, n$. Define $a_k$ via $C_k = T_{a_k}$, for all $k = 1, \ldots, 2m + 1$. 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 $c = \nu x + g, d = \mu y + g'$, for some $\nu, \mu \in \mathbb{Z}_2$ and $g, g' \in \tilde{I}$. The commutation relation $[c, d] = 0$ then implies that $\nu \mu = 0$, hence either $\nu = 0$ or $\mu = 0$. Wlog. assume that $\nu = 0$. Then, $c \in I_l$, hence $c + d \in I_l$. Thus, $c + d \in \Omega$.

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

Non-contextuality. There exists a function $\gamma|_\tilde{I} : \tilde{I} \rightarrow \mathbb{Z}_2$ that satisfies Eq. (8) on $\tilde{I}$. We now extend this function to $\Omega$ as follows. The values $\gamma(a_k)$, for $k = 1, \ldots, \xi$ 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 $\Omega$, and $d\gamma(a, b) = \beta(a, b)$ thus holds.

This establishes that the sets $\Omega$ of Eq. (13) exist for the maximum value of $\xi, \xi = 2m + 1$. One may always choose $\xi$ smaller, which neither affects closedness nor non-contextuality. □

Theorem 1 All maximal cnc sets $\Omega$ are of the form Eq. (13), with $\xi = 2m + 1$.

Proof of Theorem 1. Let $\Omega \subset E$ be closed under inference and non-contextual. We can partition the elements of $\Omega$ into two subsets, $\Omega = \{g_1, \ldots, g_\mu|g_1, \ldots, g_\mu\}$, where $\tilde{I} = \{g_1, \ldots, g_\mu\}$ are the elements of $\Omega$ which commute with the whole set. $\tilde{I}$ is an isotropic subspace since if two elements, $a$ and $b$, commute with $\Omega$, then clearly their sum,
where \( \xi \geq 2 \), the cosets \( p_1 + \tilde{I}, \ldots, p_\xi + \tilde{I} \) are distinct and \( q_1, \ldots, q_\mu \) are in the cosets \( p_1 + \tilde{I}, \ldots, p_2 + \tilde{I} \). Note that in this form, there can be no element \( p_j \) which commutes with all of \( p_1, \ldots, p_\xi \) because \( \tilde{I} \) is defined to contain all such elements. Now we consider the possible commutation relations that \( p_1, \ldots, p_\xi \) can have if \( \Omega \) is non-contextual.

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 \( \{X_1, X_2, Z_1, Z_2\} \):

\[
[p_1, p_2] = [p_1, p_4] = [p_2, p_3] = [p_3, p_4] = 0
[p_1, p_3] = [p_2, p_4] = 1
\]  
(15)

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

\[
[p_1, p_2] = [p_2, p_3] = [p_3, p_4] = 0
[p_1, p_3] = [p_1, p_4] = [p_2, p_4] = 1.
\]  
(16)

The reason is that since the set is closed under inference, it will necessarily contain the elements \( p_1 + p_2 \) and \( p_2 + p_1 \), and the elements \( p_1, p_1 + p_2, p_3 + p_4 \) have the commutation relations of Eq. (15). 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

\[
[p_1, p_2] = [p_2, p_3] = 0
[p_1, p_3] = [p_1, p_4] = [p_2, p_4] = [p_3, p_4] = 1.
\]  
(17)

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 \( p_1 + p_2, p_3 + p_4 \) have the commutation relations of Eq. (15).

To determine the possible commutation relations of the elements \( p_1, \ldots, p_\xi \), we will look at their commutativity graph \( G \). That is the undirected graph with a vertex for each of \( p_1, \ldots, p_\xi \) and an edge connecting each pair of commuting vertices. Since \( \Omega \) is non-contextual, the commutation relations of Eq. (15), Eq. (16) and Eq. (17) provide restrictions on the possible commutation relations of the elements \( p_1, \ldots, p_\xi \) of \( \Omega \). In terms of the commutativity graph \( G \), these are forbidden induced subgraphs [65].

The restriction of Eq. (15) says that \( G \) cannot have a four vertex chordless cycle \( (C_4) \) as an induced subgraph and the restriction from Eq. (16) says that \( G \) cannot have a four vertex path \( (P_4) \) as an induced subgraph. These two forbidden induced subgraphs characterize the trivially perfect graphs [59]. I.e. \( G \) must be a trivially perfect graph.

Connected trivially perfect graphs have the the property that they contain a universal vertex [60] [59]. If the commutativity graph \( G \) were connected then there would be an element \( p_j \) which commutes with all other elements of \( \{p_1, \ldots, p_\xi\} \). This is also forbidden. Therefore, the graph \( G \) is disconnected.

Given that \( G \) is disconnected, Eq. (17) provides another restriction. Namely that each connected component of \( G \) cannot have a three vertex path \( (P_3) \) as an induced subgraph. I.e. each connected component of \( G \) is a clique.

This means we can partition the elements \( \{p_1, \ldots, p_\xi\} \) into disjoint subsets

\[
\{p_1, \ldots, p_\xi\} = \{p_1, p_2, \ldots, p_\xi\} \cup \{p_2, p_3, \ldots, p_\xi\} \cup \cdots \cup \{p_\xi, 1, p_\xi, 2, \ldots, p_\xi, \xi\}
\]

where two elements commute if and only if they are in the same subset in the partition. Since the set \( \{p_1, \ldots, p_\xi\} \) 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 set 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 \( \{p_1, \ldots, p_\xi\} \) of Eq. (14) pair-wise anticommute.

Maximal cnc sets are sets of the form Eq. (14) for which \( \xi \) is maximal for a given isotropic subspace.
If the isotropic subspace \( \tilde{I} \) has dimension \( n - m \) where \( n \) is the number of qubits and \( 1 \leq m \leq n \), the maximal value of \( \xi \) is the largest number of pair-wise anticommuting Pauli operators on \( m \) qubits: \( 2m+1 \). This can be seen as follows. Consider the elements in \( E \) defined by

\[
a_k = x_k + \sum_{i=1}^{k-1} y_i, \quad b_k = z_k + \sum_{i=1}^{k-1} y_i
\]

where \( y_i = x_i + z_i \). The set \( \{a_k, b_k \mid 1 \leq k \leq n \} \) consists of pairwise anticommuting elements. There is a unique element \( c \) that would anticommute each one of the elements in this set since the set of equations

\[
[c, a_k] = [c, b_k] = 1 \quad 1 \leq k \leq n \tag{18}
\]

has a unique solution. Therefore together with this element we can construct a set of size \( 2n+1 \). We would like to show any other set of pairwise anticommuting elements whose size is \( 2n+1 \) can be mapped bijectively to the set we constructed. Suppose \( \{\tilde{a}_k, \tilde{b}_k \mid 1 \leq k \leq n \} \) is such a set. By Witt’s lemma \([20]\) the function that sends \( a_k \) to \( \tilde{a}_k \), and \( b_k \) to \( b_k \) extends to a linear map \( f : E \to E \) 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 \( \tilde{a}_k, \tilde{b}_k \), and it is given by \( f(c) \). In particular, \( 2n+1 \) is the maximal number.

Therefore, all maximal cnc sets have the form Eq. (13), with \( \xi = 2m+1 \). □

### B. 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 3.** Be \( |\text{stab} \rangle \) an \( n \)-qubit stabilizer state, with isotropic subspace \( \tilde{I} \subset E \) corresponding to its stabilizer. Then, it easily verified that \( \tilde{I} \) is non-contextual and closed under inference. Namely, \( \tilde{I} \) is of form Eq. (13), with \( m = 0, \xi = 1 \).

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

**Example 4.** Every \( n \)-qubit state of the form \( \Psi = \rho_1 \otimes |\text{stab} \rangle \langle \text{stab}|_{2\ldots n} \), with \( \rho \) 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 \( |\text{stab} \rangle_{2\ldots n} \), and three elements \( x, y, z \in E \), such that \( \tilde{T}_x = X \), \( \tilde{T}_y = Y \) and \( \tilde{T}_z = Z \). 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 \( \Omega_{xyz} := I_x \cup I_y \cup I_z \). \( \Omega_{xyz} \) is of form Eq. (13), with \( m = 1, \xi = 3 \) and \( n \geq 2 \), hence cnc by Lemma 3.

We now apply this result to the state \( \Psi = \rho_1 \otimes |\text{stab} \rangle \langle \text{stab}|_{2\ldots n} \) 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 \( |\text{stab} \rangle \langle \text{stab}| = A_1^\gamma \) (cf. Example 2). We observe that

\[
A_{\Omega_0}^{\gamma_0} \otimes A_I^\gamma = A_{\Omega_{xyz}}^\gamma,
\]

with \( \gamma := \gamma_0(a_1) + \gamma(a_{2\ldots n}) \mod 2 \). To see this, recall that \( \Omega_0 = \{0, x, y, z\} \), and note that the set \( \Omega_{xyz} \) 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 \( \Psi \) are all positively representable. Yet not all these states are mixtures of stabilizer states. Stabilizer mixedness is preserved under partial trace. Now assume that \( \Psi \) is a stabilizer mixture for all \( \rho \). Then \( \text{Tr}_{2\ldots n} \Psi = \rho_1 \) is also a stabilizer mixture. Contradiction.

We cast the combined conclusion of Examples 3 and 4 as a Lemma.

**Lemma 4** For all integers \( n \), 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 [19] and [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 Section VIII 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 \( \Omega \) we introduce the derived set \( \Omega \times a \). Denoting \( \text{Comm}(a) := \{ b \in E | [a, b] = 0 \} \) and \( \Omega_a := \Omega \cap \text{Comm}(a) \),

\[
\Omega \times a := \Omega_a \cup \{ a + b | b \in \Omega_a \}, \quad \forall a \notin \Omega.
\] (19)

Likewise, we define an update on functions \( \gamma \) invoking the measurement outcome \( s_a \) of an observable \( T_a \), namely \((\cdot) \times s_a : (\gamma : \Omega \to \mathbb{Z}_2) \mapsto (\gamma \times s_a : \Omega \times a \to \mathbb{Z}_2)\). We define this update only for \((\Omega, \gamma) \in \mathcal{V}\), and only for \( a \notin \Omega \) \([67]\). The updated function \( \gamma \times s_a : \Omega \times a \to \mathbb{Z}_2 \) is given by

\[
\gamma \times s_a(b) := \gamma(b), \quad \forall b \in \Omega_a,
\] (20a)

\[
\gamma \times s_a(b) := \gamma(a + b) + s_a + \beta(a,b), \quad \forall a + b \in \Omega_a.
\] (20b)

The rules of Eq. (20) are used to formulate the update rule for phase point operators of Eq. (2) under Pauli measurement.

Remark. Update rules similar to Eq. (20) have been used previously \([33]\) to construct a \( \psi \)-epistemic model of the multi-qubit stabilizer formalism. Those rules update the value assignments in the same way but are applied under different conditions. Specifically, the update in \([33]\) does not refer to sets \( \Omega \) satisfying the conditions of Def \([4]\).

Lemma 5 Denote the projectors \( P_a(s_a) := (I + (-1)^s_a T_a)/2 \), and be \( A^{\gamma}_{\Omega} \) a phase point operator defined through Eq. \([4]\), 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 \( A^{\gamma}_{\Omega} \) is

\[
P_a(s_a)A^{\gamma}_{\Omega}P_a(s_a) = \delta_{s_a, \gamma(a)}A^{\gamma}_{\Omega} + \frac{A^{\gamma+|a|}_{\Omega}}{2}, \quad \text{if} \ a \in \Omega,
\] (21a)

\[
P_a(s_a)A^{\gamma}_{\Omega}P_a(s_a) = \frac{1}{2}A^{\gamma\times s_a}_{\Omega}, \quad \text{if} \ a \notin \Omega.
\] (21b)

Example 2, continued. Eq. (21) entails the update of both the sets \( \Omega \) and the functions \( \gamma \). Here we only consider the former. Fig. \( \text{a} \) displays the update of the set \( \Omega \) shown in Fig. \( \text{b} \), under the measurement of (a) the observable \( X_1 \), with \( a(X_1) \notin \Omega \), and (b) the observable \( X_2 \), with \( a(X_2) \in \Omega \).

In preparation of the proof of Lemma 5 it is useful to state two relations of the function \( \beta \) for \( d = 2 \).

With the definition Eq. \([5]\) of \( \beta \) and Eq. \([3]\), the operator identities \( T_a T_a = I \) and \( T_b I = T_b \) imply that

\[
\beta(a, a) = \beta(a, 0) = \gamma(0), \quad \forall a \in E.
\] (22)

Furthermore, evaluating \( d\beta(a, a, 0) = 0 \) (see Eqs. \([6]\) and \([11]\)), and using Eq. \([22]\) yields

\[
\beta(a, b) = \beta(a, a + b), \quad \forall a, b.
\] (23)

To prove Lemma 5 we also need the following result.

Lemma 6 If \( \Omega \subset E \) is non-contextual and closed under inference, then so is \( \Omega_a \), 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 also \( c + d \in \Omega \), since \( \Omega \) is closed by assumption. Further, \([c, a] = [d, a] = 0 \) implies \([c + d, a] = 0 \), and hence \( c + d \in \Omega_a \). \( \Omega_a \) is thus closed.

Now consider non-contextuality. Since \( \Omega \) non-contextual, there exists a function \( \gamma \) such that \( d\gamma = \beta \) on \( \Omega \). Since \( \Omega_a \) is closed, \( \beta \) can be properly restricted to \( \mathcal{C}(\Omega_a) \), and so can \( \gamma \). Hence, \( d\gamma|_{\mathcal{C}(\Omega_a)} = \beta|_{\mathcal{C}(\Omega_a)} \). Thus, \( \Omega_a \) is non-contextual. \( \square \)

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

\[
I + (-1)^s_a T_a A^{\gamma}_{\Omega} I + (-1)^s_a T_a =
\]

\[
= \frac{1}{2} \sum_{b \in \Omega_a} (-1)^{\gamma(b)} T_b
\]

\[
= \frac{1}{2 \cdot 2^n} \sum_{b \in \Omega_a} (-1)^{\gamma(b)} T_b + \frac{(-1)^s_a}{2 \cdot 2^n} \sum_{b \in \Omega_a} (-1)^{\gamma(b)} T_a T_b.
\] (24)
From hereon we need to distinguish two cases, $a \in \Omega$ and $a \notin \Omega$.

\textbf{Case I:} $a \in \Omega$. Focussing on the second term in the expansion Eq. (24),

\begin{align*}
(-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(b)} T_{a+b}.
\end{align*}

Therein, in the first line we have used Eq. (5), in the second line Eq. (8), in the third line the completeness of $\Omega_a$ under inference (Lemma 6), and the fourth line is just a relabeling of the elements in $\Omega_a$. Inserting this result in the above expansion Eq. (24), we find

\begin{align*}
P_a(s_a) A_\Omega^+ P_a(s_a) &= \delta_{s_a,\gamma(a)} \frac{1}{2^a} \sum_{b \in \Omega_a} (-1)^{\gamma(b)} T_b, \tag{25}
\end{align*}

and Eq. (21a) follows.

\textbf{Case II:} $a \notin \Omega$. Substituting $b \rightarrow a+b$ in Eq. (20b) gives $\gamma \times s_a(a+b) = \gamma(b) + s_a + \beta(a,a+b)$, for $b \in \Omega_a$. With Eq. (23) we obtain

\begin{align*}
\gamma \times s_a(a+b) &= \gamma(b) + s_a + \beta(a,b), \quad \forall b \in \Omega_a. \tag{26}
\end{align*}

With this, we now look at the second term in the expansion Eq. (24),

\begin{align*}
(-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{align*}

The first line above follows with Eq. (5), and the second with Eq. (26).

Considering the first term in the expansion Eq. (24), with Eq. (20a) we have

\begin{align*}
\sum_{b \in \Omega_a} (-1)^{\gamma(b)} T_b &= \sum_{b \in \Omega_a} (-1)^{\gamma \times s_a(b)} T_b
\end{align*}

Inserting the above expressions for the two terms in Eq. (24), and using the definition Eq. (19) of $\Omega \times a$, we obtain Eq. (21b). \hfill \Box

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 new phase point operators appearing on the r.h.s. of Eq. (21) satisfy the consistency constraints of Definition 4.

\textbf{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}. \hfill \Box

\textbf{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, $p_a(s_a) := \text{Tr}(P_a(s_a)) > 0$. We have to show that under these conditions, the post-measurement state

\begin{align*}
\rho' &= \frac{P_a(s_a) \rho P_a(s_a)}{p_a(s_a)}
\end{align*}

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

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

\begin{align*}
\rho' &= \sum_{(\Omega,\gamma) \in \mathcal{V}} \frac{W_\rho(\Omega,\gamma)}{p_a(s_a)} \left( \delta_{a,\Omega} \delta_{s_a,\gamma(a)} A_\Omega^+ + A_\Omega^+ \gamma + [a, \cdot] \right) + \frac{1}{2} \delta_{a,\Omega} A_\Omega^+ \gamma \times s_a. \tag{27}
\end{align*}

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

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

(28)

The $W_\rho'(\Omega', \gamma')$ are thus linear combinations of $W_\rho(\Omega, \gamma)$ with non-negative coefficients (0 or 1/2$p_a(s_a)$). Since the $W_\rho(\Omega, \gamma)$ are non-negative by assumption, it follows that $W_\rho'(\Omega', \gamma') \geq 0$, for all $(\Omega', \gamma') \in \mathcal{V}$. \hfill \Box
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 described below in Table II outputs one sample from the joint probability distribution $p(s_{a_1}, s_{a_2}, \ldots, s_{a_N})$ of outcomes corresponding to a sequence of measurements of Pauli operators $T_{a_1}, T_{a_2}, \ldots, T_{a_N}$ ($T_{a_1}$ is measured first, $T_{a_N}$ last). If more than one sample are desired, the procedure is just repeated. We note that the observables can be chosen at runtime. 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 consider a sequence of Pauli measurements on an initial $n$-qubit state $\rho$ that is positively representable, such that the classical simulation algorithm of Table II applies. We then have the following result.

**Theorem 3** The algorithm of Table II is correct, and, if the initial probability distribution $W_\rho$ can be efficiently sampled from, it is also computationally efficient.

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 [33]. 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 mechanics. The measurement sequence can be of any length and the measurements therein may be commuting or anticommuting. In [33], 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 II also saturates this limit. To show this, we use as cnc sets $\Omega$ the six maximal isotropic subspaces of two rebits, cf. Fig. 1 b. This set of sets $\Omega$ is closed under update by Pauli measurement, as described by Eq. (21). For each such set $\Omega$, each value assignment $\gamma$ is specified by two evaluations (the other evaluations then follow via Eq. (8)). There are thus four functions $\gamma$ for each cnc set $\Omega$, hence 24 combinations in total.

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 $\rho$ with $W_\rho \geq 0$, and given the capability to sample from $W_\rho$? 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 $\rho$.

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

If sets $\Omega$ with $m = 1$ are included in the expansion, then more two-rebit states $\rho$ can be positively represented (among them, for example, $|T_1\rangle \otimes |T_2\rangle$) but on the other hand, memory consumption goes up. For $m = 1$, there are $3^2 \times 2^3 = 24$ pairs ($\Omega, \gamma$), cf. Fig. 1 h. Hence the memory consumption for configurations with $m = 1$ is $\log_2 72 \approx 6.17$ bits. (Note that the sets $\Omega$ 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 of the phase space vs. the maximum value of $m$ is displayed in Table III. The memory cost is $\log_2 |\Omega|$. The volume fraction of positively represented two-rebit and two-qubit states is displayed in Fig. 1 i in Sec-

---

**TABLE II: 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 $W_\rho$ representing the initial quantum state $\rho$.
2. For the observables $T_{a_1}, T_{a_2}, \ldots, T_{a_N}$ measured in this sequence, repeat the following steps.
   For the $i$-th measurement, set $a := a_i$.
   If $a \in \Omega$ then $\Omega$ is unchanged. Output the value $s_a = \gamma(a)$. Flip a coin.
   - if “heads” then $\gamma \rightarrow \gamma + [a, \cdot]$.
   - 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. (20).
The classical simulation algorithm of Section VI A is correct only if the quantum and the classical ways of computing \( p_a(s_a) \) and \( p' \) 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 post-measurement state separately.

**Outcome probability \( p_a(s_a) \).** Then, the quantum mechanical expression for \( p_a(s_a) \) is

\[
p_a(s_a) = \sum_{(\Omega, \gamma) \in V} W_\rho(\Omega, \gamma) \left( \delta_{a \in \Omega} \delta_{s_a, \gamma(a)} + \frac{1}{2} \sum_{a \in \Omega} \right).
\]

The classical expression \( p_a^{(c)}(s_a) \) for \( p_a(s_a) \) obtained through the algorithm of Section VI A is as follows. If \( a \in \Omega \), then the conditional probability for the outcome \( s_a \) given the state \((\Omega, \gamma)\) is \( \delta_{s_a, \gamma(a)} \). If \( a \not\in \Omega \) then the conditional probability for the outcome \( s_a \) is 1/2. Thus,

\[
p_a^{(c)}(s_a) = \sum_{(\Omega, \gamma) \in V} W_\rho(\Omega, \gamma) \left( \delta_{a \in \Omega} \delta_{s_a, \gamma(a)} + \frac{1}{2} \sum_{a \in \Omega} \right).
\]

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.

**Post-measurement state \( \rho' \).** The quantum mechanical expression for the post-measurement state \( \rho' \) has already been given in Eq. 27, 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 \( T_\alpha \) and ending up in the state \((\Omega', \gamma')\). We may invoke conditional probabilities in two ways,

\[
p((\Omega', \gamma') \cap s_a) = \sum_{(\Omega, \gamma) \in V} p((\Omega', \gamma') \cap s_a | (\Omega, \gamma)) W_\rho(\Omega, \gamma)
\]

\[
= \sum_{(\Omega, \gamma) \in V} p((\Omega', \gamma') | s_a) p_a(s_a).
\]

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

\[
W_\rho(\Omega', \gamma') = \sum_{(\Omega, \gamma) \in V} W_\rho(\Omega, \gamma) p_a((\Omega', \gamma') \cap s_a | (\Omega, \gamma)).
\]

We now infer the conditional probabilities \( p((\Omega', \gamma') \cap s_a | (\Omega, \gamma)) \) from the classical simulation algorithm of Section VI A.
\[ p((\Omega', \gamma') \cap s_a | (\Omega, \gamma)) = \begin{cases} 
\frac{1}{2} \delta_{s_a, \gamma(a)} \delta_{\Omega, \Omega} \left( \delta_{\gamma', \gamma} + \delta_{\gamma', \gamma + [a, \cdot]} \right), & \text{if } a \in \Omega, \\
\frac{1}{2} \delta_{\Omega, \Omega} \delta_{\gamma', \gamma} s_a, & \text{if } a \notin \Omega.
\end{cases} \]

Inserting this into Eq. (30), and using the resulting expression in Eq. (4), i.e. applying the map \( R \), we obtain

\[ \rho'_c = \sum_{(\Omega', \gamma') \in \mathcal{V}} \sum_{(\Omega, \gamma) \in \mathcal{V}} W_{\rho}(\Omega, \gamma) \frac{2 p_a(s_a)}{p_a(s_a)} \left( \delta_{a \in \Omega} \delta_{s_a, \gamma(a)} \delta_{\Omega, \Omega} \left( \delta_{\gamma', \gamma} + \delta_{\gamma', \gamma + [a, \cdot]} \right) + \delta_{a \in \Omega} \delta_{\Omega, \Omega} \delta_{\gamma', \gamma} s_a \right) A_{\Omega}^{\gamma'}, \]

Comparing the last expression with Eq. (27), we find that \( \rho'_c = \rho' \) for all \( a, s_a \), and the right diagram in Eq. (3) thus commutes. \( \square \)

Proof of Theorem 3. (i) Correctness. Denote by \( \rho(t) \) the state before the \( t \)-th measurement. With Lemma 5, by induction on the right diagram in Eq. (3), if \( W_{\rho(1)} \) represents the initial state \( \rho(1) \), then \( W_{\rho(t)} \) represents \( \rho(t) \) for all time steps \( t = 1, \ldots, N \). Then, by the left diagram in Eq. (3), the outcome probabilities \( p_a(s_a | s_{<t}) \), with \( s_{<t} = (s_{a_1}, \ldots, s_{a_{t-1}}) \) the measurement record prior to time \( t \), are also correct. Thus the joint outcome probability that is sampled from,

\[ p_{a_1, \ldots, a_n}(s_{a_1}, \ldots, s_{a_n}) = \prod_{t=1}^N p_a(s_a | s_{<t}), \]

is also correct.

(ii) Efficiency. We recall that all cnc sets \( \Omega \) are unions of \( O(n) \) isotropic spaces \( \Omega_i \) (Theorem 1). Further, each \( \Omega_i \) defines a stabilizer group

\[ T_{\Omega_i} := \{ T_{a} := (-1)^{\gamma(a)} T_{a} | a \in \Omega_i \}. \]

This allows us to describe \( (\Omega, \gamma) \in \mathcal{V} \) using polynomial memory by storing \( O(n) \) stabilizer tables of size \( O(n^2) \). Indeed, by Defs. 2 and Lemma 5, \( T_{\Omega_i} \) is a closed commutative group. Furthermore, with Def. 3 it holds that \( T_{a} T_{b} = T_{a+b} \gamma_b a, b \in \Omega_i \). This implies the existence of a non-trivial stabilized subspace:

\[ P_{\Omega_i} := \sum_{a \in \Omega_i} T_a / |\Omega_i| \]

is a common +1-eigenprojector of every \( T_a \in T_{\Omega_i} \), as \( T_{a} P_{\Omega_i} = \sum_{b \in \Omega_i} T_{a+b} / |\Omega_i| = \sum_{b \in \Omega_i} T_{b} / |\Omega_i| = P_{\Omega_i}, \forall T_{\Omega_i} \), which also implies \( P_{\Omega_i} = T_{\Omega} \).

We now note that the update rules in algorithm (3), namely (i) checking whether \( a \in \Omega \), (ii) evaluating \( \gamma \) on \( a \in \Omega \), (iii) updating \( \gamma \to \gamma + [a, \cdot] \), (iv) \( \Omega \to \Omega \times a \) and (v) \( \gamma \to \gamma \times s_a \), implement tasks that admit efficient classical algorithms in the stabilizer formalism [17, 54].

Rules (i) and (ii): To test \( a \in \Omega \), we check whether \( a \in \Omega, i = 1, \ldots, O(n) \). If \( a \in \Omega \) for some value of \( j \), then \( \gamma(a) \) is computed as the bit determining the phase of the stabilizer operator \( T_a^{\gamma} \in T_{\Omega_i} \). Both tasks can be solved classically efficiently via Gaussian elimination given the stabilizer table data [17, 54]. Rule (iii): \( \gamma \) is updated to \( \gamma' = \gamma + [a, \cdot] \) (classically efficiently) evaluating \( \gamma(\cdot) + [a, \cdot] \) on the generators of every \( \Omega_i \). Rules (iv) and (v): For all \( j, T_{\Omega_i}^{\gamma(a_1), s_a} \) is the stabilizer group resulting from the measurement of \( T_a \) with outcome \( s_a \) on a state with stabilizer group \( T_{\Omega_i}^{\gamma}. \) This update can be efficiently performed using the standard measurement update-rule of Ref. [17, 54] to every stabilizer table in the description of \( (\Omega, \gamma) \). Thus, all steps of the algorithm run in polynomial time. \( \square \)

VII. 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 non-contextual 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].

This situation admits two interpretations: (a) The contextuality of the HVM in \( d = 2 \) remains a fundamental difference to the case of odd \( d \), where at least some quantum states admit a non-contextual HVM. (b) Contextuality, i.e., the unviability of non-contextual HVMs, is not a sufficiently stringent 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.
Below we characterize the present HVM; namely in Sec.
 we describe the structure of its value assignments, and in Sec.
 we provide our interpretation of the present HVM: it should be considered classical even though it is contextual.

A. Description of the HVM

Our HVM consists of a triple \((\Lambda, \{h^\lambda\}, p_\lambda)\) where \(\Lambda = \mathcal{V}\) or \(\mathcal{V}_M\), \(h^\lambda\) is a compatible family of distributions on the set of outcomes on contexts and \(p_\lambda\) is a probability distribution on the set \(\Lambda\) of hidden variables.

For each \(\alpha = (\Omega, \gamma)\) we define \(h^\alpha\) by

\[
h^\alpha_I(s) = \text{Tr}(P_s A^\alpha_{i_1}^\alpha).
\]

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 \(h^\alpha(s)\) in their explicit form.

Lemma 9 Let \(P_s\) denote the projector corresponding to the non-contextual value assignment \(s : I \rightarrow \mathbb{Z}_2\). Then we have

\[
h^\alpha_I(\Omega, \gamma)(s) = \frac{|I \cap \Omega|}{|I|} \delta_{s|_\Omega, \gamma|_\Omega}.
\]

From Eq. (33) 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 we observe that the \(\{h^\alpha_I\}\) form compatible families,

\[
h^\alpha_I|_{I \cap I'} = h^\alpha_{I'}|_{I \cap I'}, \forall I, I', \forall \alpha \in \mathcal{V}.
\]

The HVM of Eq. (32) is (a) contextual in the sense of Abramsky and Brandenburger \cite{AB}, (b) preparation and transformation contextual, as well as measurement-non-contextual, in the sense of Spekkens \cite{Sp}, and (c) contextual for sequences of transformations in the recently-introduced sense of Mansfield and Kashefi \cite{MK}.

B. Relation to other models and frameworks

1. Sheaf-theoretic formulation

The HVM described by Eq. (32) may be discussed in the sheaf theoretic language of Abramsky and Brandenburger \cite{AB}. We will write \(\mathcal{E}(\Omega)\) for the set of value assignments defined on a subset \(\Omega \subset E\). Let us start with the case where the generalized phase space \(\mathcal{V}_M\) coincides with global value assignments on the whole space \(E\). Given an isotropic subspace \(I\) we can restrict global value assignments

\[
\text{res}_I : \mathcal{V}_M \rightarrow \mathcal{E}(I).
\]

Writing \(D(X)\) for the set of \(\mathbb{R}_{\geq 0}\)-distributions on a set \(X\) we can think of a ncHVM element \(h^\lambda\) associated to a global value assignment \(\lambda\) as an element of \(DV_M\). The odd prime qudit case is an example of this sort since \(\mathcal{V}_M\) coincides with the set of global value assignments.

Our present HVM diverges from this picture in two ways (1) \(\mathcal{V}_M\) consists of partial value assignments, (2) the analog of the restriction in Eq. (34) gives a uniform probability distribution outside the partial value assignment. Therefore the target of our restriction map has to be \(\mathcal{E}(I)\), and we need to specify how to restrict a partial value assignment \((\Omega, \gamma)\). We define

\[
\text{ures}_I : \mathcal{V}_M \rightarrow \mathcal{E}(I)
\]

by the average of delta distributions

\[
\text{ures}_I(\Omega, \gamma) = \frac{|I \cap \Omega|}{|I|} \sum_{\lambda: I = I} \delta_{\lambda}.
\]

where the summation runs over value assignments \(\lambda : I \rightarrow \mathbb{Z}_2\) that matches with \(\gamma\) over \(I \cap \Omega\). This formula is consistent with Eq. (33) in the sense that

\[
\text{ures}(\xi) = h^\Omega(\xi).
\]

2. The eight state model and its extensions

As explained in section \(\text{III.D}\) example 1, our model extends the eight-state hidden variable model for \(\mathcal{V}_2\), which can also positively represent the single-qubit stabilizer subtheory as well as all single-qubit quantum states. Furthermore, it strictly extends a recently proposed epistemic model for the multiqubit stabilizer subtheory \cite{KRM}: the latter positively represents mixtures of stabilizer states, while our model can also positively describe states outside the stabilizer polytope (section \(\text{IV.B}\)).

3. Ontological models for quantum circuits

Both our model as well as the Gottesman-Knill simulation fit within a family of ontological models for adaptive quantum circuits (including Clifford ones) considered in \cite{FB}. Therein, a \(\Omega(n^2)\) lower bound was given for the memory needed to simulate the multiqubit stabilizer subtheory. This bound is fulfilled by our simulation method in section \(\text{VI.D}\).
4. Spekkens’ contextuality

In Ref. [60] Spekkens introduced a generalized framework to study contextuality in ontological models based on a notion of operational equivalence. Therein, he distinguished three kinds of contextuality for preparations, measurements and transformations. Though we will not revisit these concepts, it is easy to see that our model is contextual for preparations and transformations in Spekkens’ sense, though measurements are non-contextual (hence, “classical”). First, it follows easily from the linearity of Eq. (32) that our model is measurement non-contextual. This requires considering probabilistic assignments for projective measurements (or “outcome indeterministic assignments for sharp measurements” in the notation of [60]). Second, our model inherits the transformation contextuality for unitary mixtures of the eight-state model [36]. in particular, the fully depolarizing channel admits multiple ontic representations by an argument analogous to Eq. (5-7)]. Last, our model is preparation contextual. For example, all stabilizer states admit multiple decompositions, namely in terms of sets Ω with m = 0, or with m = 1.

The above properties are reminiscent of Beltrametti and Bugajski [61] for Quantum Theory, which is measurement non-contextual but preparation and contextual [60]. Yet, our model has more classical features than Beltrametti-Bugajski’s, namely, it is effiencly classically simulable, it has a matching Wigner function, and, as discussed in Appendix D, it is ψ-epistemic.

5. Sequential contextuality

Ref. [52] recently introduced a new notion of contextuality for sequences of transformations. Therein, the authors noted that the constant-depth quantum circuits of [51] are sequential-transformation contextual. We point out that, because the circuits of Ref. [51] consist Clifford operations and Pauli measurements only, it follows that our model is sequential-transformation contextual.

C. Interpretation: coexistence of classicality and contextuality

We argue that, even though the HVM of Eq. (32) is contextual, as implied by established definitions [37], it should be considered as classical. The reason is the following. The only resource the HVM of Eq. (32) uses beyond those required by non-contextual HVMs with deterministic value assignments is that of classical uniform randomness (in the evaluation of value assignments). Such use of classical randomness should not render the present HVM genuinely quantum.

We denote the HVM of Eq. (32) as “HVM with partial non-contextual value assignments” (pnc HVM), and propose it for further investigation. It has the following properties:

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

(II) It is conceptually very close to a non-contextual HVM with deterministic value assignments. The only difference is that for every state of the HVM, only a subset Ω of the observables has deterministic values, and all other observables are evaluated by coin flipping.

(III) It satisfies the requirements (i) and (ii) put forth in the introduction of this section; namely (III-i) It applies to all QCM on states with positive W. (III-ii) The pnc HVM reduces to the standard non-contextual HVM with deterministic value assignments for QCM in odd d.

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

As we have established in the previous sections, $W_\rho < 0$ is a precondition for quantum speedup. For this regime, we can quantify the cost of classical simulation by a phase space method, and establish a monotone for state magic.

A. Robustness

In close analogy to the “robustness of magic” [24] $R_S$ (the subscript S is for “stabilizer”), we define a phase space robustness $R$, through

$$R(\rho) := \min_{W} \|W\|_1,$$  \hspace{1cm} (36)

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

Since the definitions of the robustness $R$ and of the robustness of magic $R_S$ [24] 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 10 For all quantum states $\rho$, of any number n of qubits, the phase space robustness $R(\rho)$ and the robustness of magic $R_S(\rho)$ are related via

$$R(\rho) \leq R_S(\rho) \leq (4n + 1) R(\rho).$$  \hspace{1cm} (37)

Thus, the phase space robustness $R$ is never larger than the robustness of magic, but can only be moderately smaller. The proof of Lemma 10 is given in Appendix C.
Thus, the robustness \( R \) respectively represented by \( W \) sources are quantum states which cannot be positively represented states which are not mixtures of stabilizer states. Such states are called (iv) bound magic states. We have seen an example of them in Section IV B, the general 1-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 only be converted into other bound states or into free states by the free operations, but never into a resource.

The question of inter-convertibility 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 \( R \), defined in Eq. (36) and already known to measure hardness of classical simulation by sampling, is a monotone.

**Theorem 4** The robustness \( R \) 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. Be \( \mathrm{Cl}_n \) the \( n \)-qubit Clifford group. It acts on the \( n \)-qubit Pauli operators via

\[
h(T_a) := hT_a^{-1} = (-1)^{\Phi_h(a)}T^\dagger_{ha}, \ \forall h \in \mathrm{Cl}_n.
\]

This implies an action of the Clifford group on the phase point operators \( A^\dagger_h \), which in turn induces an action on the sets \( \Omega \) and the functions \( \gamma \), via

\[
h(A^\dagger_h) = \frac{1}{2^n} \sum_{a \in \Omega} (-1)^{\gamma(a)}h(T_a) = \frac{1}{2^n} \sum_{b \in \Omega'} (-1)^{\gamma'(b)}T^\dagger_b.
\]

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

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

Henceforth we denote \( \Omega' \) as \( h \cdot \Omega \) and \( \gamma' \) as \( h \cdot \gamma \), to emphasize the dependence on \( h \in \mathrm{Cl}_n \).

For use in the proof below we quote Lemma 3 from \[38\] which says that, for any face \((a,b) \in \Omega \times \Omega\),

\[
\Phi_h(\partial(a,b)) = \beta(ha, hb) + \beta(a,b) \mod 2.
\]

We then have the following result.

**Lemma 11** \( V \) is mapped to itself under \( \mathrm{Cl}_n \), and the quasiprobability function \( W \) transforms covariantly. That is, if the state \( \rho \) can be described by \( W_\rho \) through Eq. (3), then for any \( h \in \mathrm{Cl}_n \) the state \( h\rho h^{-1} \) can be described by a quasiprobability function \( W_{h\rho h^{-1}} \) defined by

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

**Remark 3:** We say “the state \( \rho \) can be described by \( W_\rho \)” rather than “is described” because \( W_\rho \) is not unique.

**Proof of Lemma 11** First, we show that the phase space \( V \) is closed under the action of \( \mathrm{Cl}_n \), i.e., if \( (\Omega, \gamma) \in V \) then \( (\Omega', \gamma') \in V \). 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 $a + b \in \Omega \leq$ the assumption of closedness. Hence $\Omega'$ is closed under inference.

(iii) $\gamma'$ satisfies Eq. [3]. With the definition of $\gamma'$ we have (all addition mod 2)

$$d\gamma'(ha, hb) = d\gamma(a, b) + \Phi_h(\partial(a, b)) = d\gamma(a, b) + \beta(ha, hb) + \beta(a, b) = \beta(ha, hb).$$

Therein, in the second line we have used Eq. [3]. Thus, $\gamma'$ satisfies Eq. [3] on its domain.

(ii) $\Omega'$ is non-contextual. With $\gamma'$ we have just proved the existence of a function on $\Omega'$ that satisfies Eq. [3].

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

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

Next we turn to the covariance of $W$ under $\mathrm{Cl}_n$.

We have

$$h \rho h^\dagger = \sum_{(\Omega, \gamma) \in \mathcal{V}} W_\rho(\Omega, \gamma) h(A^\gamma_{\Omega}).$$

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

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

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

We are now ready to prove the monotonicity of $\Re$, as stated in Theorem [4].

Proof of Theorem [4] (a) Clifford unitaries. With Lemma [11] we have that for any $n$-qubit Clifford gate $h$ applied to any $n$-qubit state $\rho$, the quasiprobability distribution $W_{h \rho h^\dagger}$ can be related to $W_\rho$ via the covariance condition Eq. [38]. Since $W$ is non-unique, there may a priori be a representation $W_{h \rho h^\dagger}$ with smaller 1-norm, and thus it holds that Eq. [39].

(b) Pauli measurements. We consider the measurement of a Pauli observable $T_a$ on a quantum state $\rho$. Denote by $\rho_{a,s_a}$ the normalized post-measurement states for the outcomes $s_a = 0, 1$, respectively. We have to show that, for all $n$, for all $a \in \mathbb{Z}_2^2 \times \mathbb{Z}_2^n$ and all $n$-qubit states $\rho$ it holds that

$$p_a(0) \Re(\rho_{a,0}) + p_a(1) \Re(\rho_{a,1}) \leq \Re(\rho).$$

With Eq. [25], we can write $p_a(0) W_{\rho_{a,0}} = W_+ + W_0 + W_-$, and $p_a(1) W_{\rho_{a,1}} = W_0 + W_+$ where

$$W_+(\Omega', \gamma') := \sum_{(\Omega, \gamma) \in \mathcal{V}} W_\rho(\Omega, \gamma) \frac{1}{2} \delta_{a \in \Omega} \delta_{\gamma(0)} \left( \delta(\Omega', \gamma'), (\Omega, \gamma) + \delta(\Omega', \gamma'), (\Omega, \gamma + [a, \gamma]) \right),$$

$$W_-(\Omega', \gamma') := \sum_{(\Omega, \gamma) \in \mathcal{V}} W_\rho(\Omega, \gamma) \frac{1}{2} \delta_{a \in \Omega} \delta_{\gamma(1)} \left( \delta(\Omega', \gamma'), (\Omega, \gamma) + \delta(\Omega', \gamma'), (\Omega, \gamma + [a, \gamma]) \right),$$

$$W_0(\Omega', \gamma') := \sum_{(\Omega, \gamma) \in \mathcal{V}} W_\rho(\Omega, \gamma) \frac{1}{2} \delta_{a \in \Omega} \delta_{\gamma} \left( \delta(\Omega', \gamma'), (\Omega, \gamma \cdot s_a = 0) \right),$$

$$W_0(\Omega', \gamma') := \sum_{(\Omega, \gamma) \in \mathcal{V}} W_\rho(\Omega, \gamma) \frac{1}{2} \delta_{a \in \Omega} \delta_{\gamma} \left( \delta(\Omega', \gamma'), (\Omega, \gamma \cdot s_a = 1) \right).$$

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

$$p_{a,0} \Re(\rho_{a,0}) + p_{a,1} \Re(\rho_{a,1}) \leq ||W_+||_1 + ||W_-||_1 + ||W_+||_1 + ||W_-||_1 = ||W_+||_1 + ||W_-||_1.$$
TABLE IV: Robustness values of selected magic states. For robustness of magic ($R_S$), also see [56].

| State | $R$  | $R_S$   |
|-------|------|---------|
| $|H\rangle^\otimes d$ | 1.0  | 1.7472  |
| $|T\rangle^\otimes d$ | 1.0  | 2.23205 |
| $|H\rangle^\otimes 3$ | 1.283 | 2.189   |
| $|T\rangle^\otimes 3$ | 1.385 | 3.09807 |
| Hoggar | 1.80  | 3.8000  |

Inserting the last two relations into Ineq. (42), we arrive at

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

Since $R(\rho) = \|W_\rho\|_1$ by assumption, Eq. (40) follows. □

D. Numerical results

In Table IV and Fig. 7 we present numerical values [60] for the robustness of key magic states, and compare them to robustness of magic as defined by Howard and Campbell [24]. Table IV summarizes the robustness comparisons for the common magic states, as well as the maximal-robustness Hoggar state [16]. 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 $R$ 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 $W(\rho)$ w.r.t. Eq. (40) for a given $n$-qubit state $\rho$ always was non-zero only on $4^n$ phase space points, or fewer. $4^n$ is only a tiny fraction of the whole phase space $V$, 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 $\rho$. We can now explain the initially puzzling upper bound on the size of the support, $4^n$.

The phase space robustness $R$ of a state $\rho$ defined in Eq. (36) is the solution to the convex optimization problem

$$\min_q \left\{ \|q\|_1 : Mq = b, \right\},$$  

(43)

where $M_{i,j} = \text{Tr}(A_{\alpha_j}P_i)$, $b_i = \text{Tr}(\rho P_i)$, $\{\alpha_j : 1 \leq j \leq |V|\}$ is an enumeration of the phase points and $P_i$ are the $n$-qubit Pauli operators. For each variable $q_j$ in Eq. (43), define two new variables $q^+_j := \max(0, q_j)$ and $q^-_j := \max(0, -q_j)$. Then the convex optimization problem of Eq. (43) is equivalent to the standard form linear program

$$\min_{\tilde{q}} \left\{ \sum_j q^+_j + q^-_j : \tilde{M}\tilde{q} = b, \tilde{q} \geq 0 \right\},$$  

(44)

where $\tilde{M} = \begin{bmatrix} M & -M \end{bmatrix}$ and $\tilde{q} = \begin{bmatrix} q^+T & q^-T \end{bmatrix}^T$. This doubles the number of variables but does not change the number of 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 state can have robustness less than 1), by the fundamental theorem of linear programming, for any physical state, Eq. (44) has a solution at a vertex of the feasible polytope [58].

Since Eq. (44) has a constraint for each $n$-qubit Pauli operator (including the identity), this means any state $\rho$ has a robustness-minimizing expansion in phase point operators with no more than $4^n$ non-zero coefficients.

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. When $d$ is odd, then $W$ reduces to the familiar Wigner function for finite-dimensional systems defined by Gross [29]. However, more importantly, also for the case of qubits ($d = 2$), $W$ 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) based on sampling.

Once this fundamental property is established, it is natural to turn to the efficiency (or non-efficiency) of classical simulation in the various regimes, and to 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 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 \( W_p \) is contextual, as a consequence of Mermin’s square.

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

In this regard, we have argued that contextuality in quantum systems is not sufficient as an indicator of genuine quantumness. We have proposed the notion of “HVM with partial non-contextual value assignments” in which classicality and contextuality coexist.

Acknowledgments. We thank Piers Lillystone (JVB, CO, RR, ET) and Shane Mansfield (JVB) for discussion. This work is funded by NSERC (CO, RR, ET, MZ), Cifar (RR) and ERC (JBV).

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 \( A^\gamma_{\Omega} \) 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

\[
\begin{align}
\eta|_\Omega &= 0, \\
\eta|_{\tilde{\Omega}} &= 0.
\end{align}
\]

The condition of Eq. (A1a) need only be satisfied 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

\[
A^\gamma_{\Omega} = \frac{1}{|\tilde{\Omega}|} \sum_{\tilde{\gamma} \in \tilde{\Gamma}} A^\tilde{\gamma}_{\tilde{\Omega}}, \tag{A2}
\]

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).

\[
\frac{1}{|\tilde{\Omega}|} \sum_{\tilde{\gamma} \in \tilde{\Gamma}} A^\tilde{\gamma}_{\tilde{\Omega}} = \frac{1}{|\tilde{\Omega}|} \sum_{\tilde{\gamma} \in \tilde{\Gamma}} \frac{1}{d^n} \sum_{a \in \tilde{\Omega}} \omega^{\tilde{\gamma}(a)} T_a = \frac{1}{|\tilde{\Omega}|} \frac{1}{d^n} \sum_{a \in \tilde{\Omega}} \omega^{\tilde{\gamma}(a)} T_a \sum_{\eta \in U} \omega^{\eta(a)}.
\]

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

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

Furthermore, note \( |\tilde{\Omega}| = |U| \).

(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

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

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

\[
\frac{1}{|\tilde{\Omega}|} \sum_{\tilde{\gamma} \in \tilde{\Gamma}} A^\tilde{\gamma}_{\tilde{\Omega}} = \frac{1}{d^n} \sum_{a \in \Omega} \omega^{\gamma(a)} T_a = A^\gamma_{\Omega}.
\]

This proves Eq. (A2). Now, wlog. we may choose \( \tilde{\Omega} \) to be maximal. Since by definition any set \( \Omega \) 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 \( \mathcal{V}_M \).

\[
\rho = \sum_{\Omega, \gamma} c(\Omega, \gamma) A^\gamma_{\Omega} = \sum_{\Omega, \gamma} \frac{c(\Omega, \gamma)}{|\tilde{\Omega}(\Omega)|} \sum_{\tilde{\gamma} \in \tilde{\Gamma}_{\Omega(\Omega)}} A^\tilde{\gamma}_{\tilde{\Omega}(\Omega)}.
\]

If the expansion coefficients on the l.h.s. are positive, so they are on the r.h.s. \( \square \)
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 \(\Omega\) doesn’t change, and we only need to check the properties in Def. 4 that concern the function update, i.e., Eqs. (8), (3).

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

\[
d(\gamma + [a, \cdot])(f) = d\gamma(f) + [a, \cdot](\partial f) = d\gamma(f) + [a, c] + [a, d] + [a, c + d]
\]

Thus, \(\gamma + [a, \cdot]\) satisfies Eq. (8).

Furthermore, assume that \(\gamma\) satisfies Eq. (3). Then, \((\gamma + [a, \cdot])(0) = \gamma(0) + [a, 0] = \gamma(a)\). 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 Def. 4, namely (I) \(\Omega \times a\) is closed under inference, (II) \(\Omega \times a\) is non-contextual, (III) \(\gamma \times s_a\) satisfies Eq. (8), and (IV) \(\gamma \times s_a\) satisfies Eq. (3).

(I): Consider \(c, d, d' \in \Omega \times a\), with \([c, d] = 0\), and denote \(c' = c + a, d' = d + a\). There are three sub-cases. (ii) \(c, d \in \Omega_a\). Then, \(c + d \in \Omega_a\), since \(\Omega_a\) 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 \(\Omega \times a\), \(d' \notin \Omega_a\). Thus, \(c + d = c + (d' + a) = (c + d') + a\). Now, since \([c, d] = 0\), the assumption \([a, c] = 0\) and \([c, d'] = 0\), and hence \(\Omega_a\) is closed by Lemma 6. Then, \(c + d \in \Omega_a\) by construction of \(\Omega \times a\), w. \(c + d \in \Omega \times a\).

(iii) \(c, d \notin \Omega_a\). By construction of \(\Omega \times a\), \(c', d' \notin \Omega_a\). Thus, \(c + d = (c' + a) + (d' + a) = (c' + d')\). Further \(c' + d' = 0\). Since \(\Omega_a\) 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.

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

(i) \(c, d, c + d \in \Omega_a\). Since \(\Omega_a \subset \Omega\) and with Eq. (20a), 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 \(\partial f\) consisting of the edges \(c, d\) and \(c + d\),

\[
d(\gamma \times s_a)(f) = \gamma \times s_a(c) + \gamma \times s_a(d) + \gamma \times s_a(c + d)
\]

\[
= \gamma(c) + \gamma(d') + \gamma(c + d') + \beta(a, d) + \beta(a, c + d)
\]

\[
= \beta(c, d') + \beta(a, d) + \beta(a, c + d)
\]

\[
= \beta(c, d).
\]

Therein, in the second line we have used Eq. (20), in the third line Eq. (5), in the fourth line Eq. (23), and in the fifth line \(d\beta(a, d, c) = 0\), cf. Eqs. (9) and (11).

(II): Per Def. 3, \(\Omega \times a\) is non-contextual if there is a function \(\Gamma : \Omega \times a \to \mathbb{Z}_2\) that satisfies \(d\Gamma = \beta\). We have explicitly constructed such a function in (III) above, \(\Gamma := \gamma \times s_a\).

(IV): Assume that \(\gamma : \Omega \to \mathbb{Z}_2\) satisfies Eq. (3). Since \(0 \in \Omega_a\) for all cncl sets \(\Omega\), with Eq. (20a) 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 10

Proof of Lemma 10. Recall from Lemma 3 that each set \(\Omega\) can be written in the form \(\Omega = \bigcup_{k=1}^{m} I_k\), where each \(I_k\) is an isotropic subspace, \(I_k = \langle a_k, I \rangle\), \(a_k \in E\). Therefore, for all \((\Omega, \gamma) \in \mathcal{V}\), it holds that

\[
A^\gamma_\Omega = \left( \sum_{k=1}^{m} A^\gamma_{I_k} \right) - (\xi(\Omega) - 1)A^\gamma_\Omega|_{\Gamma}.
\]

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 \(A^\gamma_\Omega\) can thus be expanded as

\[
\delta(\Omega, \gamma) = \left( \sum_{k=1}^{m} \delta(I_k, \gamma|_{I_k}) \right) - (2 \xi(\Omega) - 1)\delta(\Omega|_{\Gamma}).
\]

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

\[
\|\delta(\Omega, \gamma)\|_{1, S} \leq 4n + 1.
\]

(C1)

Now, for any given state \(\rho\) consider the optimal representation \(W_\rho\), i.e., the one with minimal norm
The stabilizer states induce an expansion in terms of type \( m \) stabilizer states, corresponding to phase point operators coefficients. 

Thus, an expansion in terms of stabilizer states induces an expansion in terms of phase point operators \( A_{ij}^{\Omega} \), with the same non-zero coefficients. \( \Box \)

**Appendix D: Other notions of quantumness:**

\( \psi \)-onticity

Beyond contextuality, \( \psi \)-onticity (understood as the absence of a \( \psi \)-epistemic description) has been singled out as a potential genuine feature of Quantum Theory [48] [49]. In \( \psi \)-epistemic models, pure states can be interpreted as “states of knowledge”, while in \( \psi \)-ontic models, pure states are “real”, in the sense that they can be uniquely identified from any ontic-state in the support of its hidden variable representation (which can be interpreted as an “element of reality”). For instance, the wavefunction defines a \( \psi \)-ontic [39] (preparation contextual [60]) description of Quantum Theory [61]. This illustrates the huge descriptive power of \( \psi \)-ontic models.

Recently, Ref. [35] has shown that the multi-qubit stabilizer formalism admits a \( \psi \)-epistemic model. The latter model is discussed therein to be more “classical” than the Gottesman-Knill simulation, as the latter is \( \psi \)-ontic: stabilizer states are uniquely identified by a stabilizer group, which are the ontic states. Our new hidden variable model is \( \psi \)-epistemic, like the model in [35], but has the additional feature of being able to describe bound magic states. To illustrate this, note that for \( n = 1 \), the pure non-orthogonal stabilizer states \((I + X)/2, (I + Z)/2\) can be written as mixtures of point operators with overlapping support (thus are “ontologically indistinct” [49]):

\[
\begin{align*}
\frac{I + X}{2} &= \sum_{b,c \in \{\pm 1\}} \frac{1}{4} I + X + bY + cZ, \\
\frac{I + Z}{2} &= \sum_{d,e \in \{\pm 1\}} \frac{1}{4} I + dX + eY + Z,
\end{align*}
\]

where averaging over \( b, c, d, e \) cancels undesired Pauli components. This makes the model \( \psi \)-epistemic [35] [49]. Using the construction in in section [IVB] Example 4, it follows by a similar argument that any pair of \( n \)-qubit pure stabilizer states of the form \( \rho_1 \otimes |\text{stab}\rangle_{2,...,n} \), \( \rho_1 \in \{ (I + X)/2, (I + Z)/2\} \) is ontologically indistinct: specifically, they can be represented as overlapping mixtures over a fixed set of “majorana orbits”. Since all pure stabilizer states are equivalent via a Clifford unitary, it follows that every pure stabilizer state in our model is ontologically indistinct from several other pure stabilizer states, which is a stronger requirement than bare psi-epistemicity [35] [49]. (In fact, the number of such states grows at least linearly in \( n \)). Similarly to [35], this technique also allows us to represent \( n \)-qubit stabilizer states storing the information of only \( n - 1 \) stabilizers.

---

[1] P. Ehrenfest, *Bemerkung über die angenäherte Gültigkeit der klassischen Mechanik innerhalb der Quantenmechanik*, Z. Physik 45, 455 (1927).

[2] A. Einstein, B. Podolsky; 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*, Int. J. Theor. Phys. 21, 467 (1982).

[5] E. Bernstein and U. Vazirani. *Quantum complexity theory*. SIAM J. Comput., 26(5):14111473, (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
[50] Shane Mansfield, Reality of the quantum state: Towards a stronger $\psi$-ontology theorem, Phys. Rev. A 94, 042124 (2016).

[51] Sergey Bravyi, David Gosset, Robert Koenig, Quantum advantage with shallow circuits, Science Vol. 362, Issue 6412, pp. 308-311 (2018).

[52] Shane Mansfield and Elham Kashefi, Quantum Advantage from Sequential-Transformation Contextuality, Phys. Rev. Lett. 121, 230401 (2018).

[53] J. Bermejo-Vega and M. Van Den Nest, Classical simulations of Abelian-group normalizer circuits with intermediate measurements. Quantum Info. Comput. 14, 3-4, 181-216 (2014).

[54] D. Gottesman, The Heisenberg representation of quantum computers: Proceedings of the XXII International Colloquium on Group Theoretical Methods in Physics. International Press, 1999.

[55] Markus Heinrich, David Gross, Robustness of Magic and Symmetries of the Stabiliser Polytope, Preprint. arXiv:1807.10296 [quant-ph].

[56] Fernando G.S.L. Brandao and Gilad Gour, Reversible Framework for Quantum Resource Theories, Phys. Rev. Lett. 115, 070503 (2015).

[57] J. Nocedal and S. Wright, Numerical Optimization. Springer Science and Business Media, 2 edition, 2000.

[58] M. C. Golumbic, Trivially perfect graphs, Discrete Mathematics 24 (1) 105-107 (1978).

[59] R. W. Spekkens, Contextuality for preparations, transformations, and unsharp measurements Phys. Rev. A 71, 052108 (2005).

[60] E. G. Beltrametti and S Bugajski, A classical extension of quantum mechanics, Journal of Physics A: Mathematical and General, Volume 28, Number 12 (1995).

[61] Joshua B. Ruebeck, Piers Lillystone, and Joseph Emerson, Epistemic interpretations of quantum theory have a measurement problem, arXiv:1812.08218.

[62] Patrick Rall, Daniel Liang, Jeremy Cook and William Kretschmer. Simulation of Qubit Quantum Circuits via Pauli Propagation, 2019; arXiv:1901.09070 [quant-ph].

[63] Aschbacher, Michael. Finite group theory, Vol. 10. Cambridge University Press, 2000.

[64] In fact, we have equality, $R(h\rho h^\dagger) = R(\rho)$, since unitary gates are reversible, but we do not need this slightly stronger statement for the present proof.

[65] Our calculations use the software packages CVXPY [42] and Gurobi [43].