# Classical simulation of noncontextual Pauli Hamiltonians

William M. Kirby and Peter J. Love*
*Department of Physics and Astronomy, Tufts University, Medford, Massachusetts 02155, USA*

(Received 20 February 2020; accepted 19 August 2020; published 24 September 2020)

Noncontextual Pauli Hamiltonians decompose into sets of Pauli terms to which joint values may be assigned without contradiction. We construct a quasiquantized model for noncontextual Pauli Hamiltonians. Using this model, we give an algorithm to classically simulate the noncontextual variational quantum eigensolver. We also use the model to show that the noncontextual Hamiltonian problem is NP-complete. Finally, we explore the applicability of our quasiquantized model as an approximate simulation tool for contextual Hamiltonians. These results support the notion of noncontextuality as classicality in near-term quantum algorithms.

DOI: 10.1103/PhysRevA.102.032418

## I. INTRODUCTION

Simulation of quantum systems using the *variational quantum eigensolver* (VQE) is a promising application for noisy intermediate-scale quantum (NISQ) computers [1,2]. VQE is advantageous in the NISQ era because the necessary circuit depths are small compared with other quantum simulation methods, such as phase estimation [2]. Small-scale VQE experiments have been performed in a variety of qubit architectures to simulate systems from molecular to high-energy physics [2–16].

In this paper we give a classical simulation technique for *noncontextual Hamiltonians*, defined in our previous work [17]. Contextuality is an indicator of nonclassicality [18–20] that is related to negativity of quasiprobability representations [21–26] and resources for quantum computation [27–35]. Recent work has led to a variety of approaches to characterizing and quantifying contextuality [17,36–48].

In the present work we show that when a Hamiltonian is noncontextual according to the criterion of [17], we may construct a *quasiquantized* classical model of the associated VQE instance. Quasiquantized models are descriptions of classical subtheories of quantum mechanics, such as the stabilizer subtheory for odd-dimensional qudits, and Gaussian subtheory for continuous systems [21,49–52].

Our quasiquantized model is a hidden-variable theory for prepare-and-measure scenarios where all states are allowed but the measurements must correspond to a noncontextual set of Pauli observables. This contrasts with [47], in which it is shown that for noncontextual states (in the sense of being positively representable over a noncontextual phase space), any set of Pauli measurements permits classical description.

The problem of approximating the ground state energy of a classical or quantum Hamiltonian can be NP-complete or QMA-complete, respectively; NP-complete (QMA-complete) problems are the hardest decision problems whose answers can be efficiently verified on classical (quantum) computers [53–64]. Our quasiquantized model allows us to show that the noncontextual Hamiltonian problem (see Sec. I B) is only NP-complete, rather than QMA-complete [54]. In other words, the fact that we can describe the VQE procedure in terms of nonnegative joint probabilities places noncontextual VQE unambiguously in the realm of the classical, and this result extends to the computational complexity of the problem.

In Sec. II we show how to construct a quasiquantized model for any noncontextual Hamiltonian. The states of the quasiquantized model are probability distributions that correspond to quantum states. We prove that these probability distributions reproduce expectation values for the Hamiltonian terms, including the expectation values corresponding to Hamiltonian eigenstates. In Sec. III we use our model to construct a classical simulation algorithm for VQE. We also show that the noncontextual Hamiltonian problem is NP-complete. In Sec. IV we show how to approximate contextual Hamiltonians by noncontextual Hamiltonians. In some cases we reach chemical accuracy using this approximation, and in most cases we outperform existing experimental results. In Sec. V we summarize and discuss our results.

### A. Variational quantum eigensolvers

In VQE we wish to minimize the energy expectation value of a Hamiltonian $H$:

$$\langle H\rangle = \sum_{P\in\mathcal{S}} h_P \langle P\rangle, \tag{1}$$

where the $h_P$ are real coefficients, and $\mathcal{S}$ is the set of Pauli operators in the support of the Hamiltonian (see [65] for a review of VQE). The energy expectation value (1) is estimated by preparing a physical ansatz on a quantum device, and evaluating the expectation value of each Pauli term $P \in \mathcal{S}$ separately. The weighted sum (1) is then treated as an objective function for a classical optimization of the ansatz parameters [2].

### B. Local Hamiltonian problems

The $k$-local Hamiltonian problem is the decision problem of whether the ground state energy of a $k$-local Hamiltonian lies below some specified energy gap [with size at least

*Also at Brookhaven National Laboratory.

 

$1/\mathrm{poly}(n)$, for $n$ qubits], given the promise that the ground state does not lie within the gap [54]. The $k$-local Hamiltonian problem is QMA-complete, and has also been studied for various subsets of Hamiltonians (such as $k$-local commuting Hamiltonians, and Hamiltonians with specific interaction graphs) [53–64]. The complexity of the $k$-local commuting Hamiltonian problem is a long-standing open problem [57,61–64].

In this paper we focus on Pauli Hamiltonians. The complexity of the corresponding commuting problem is known: the $k$-local commuting Pauli Hamiltonian problem is in NP, since it is a special case of the $k$-local commuting, factorizable qubit Hamiltonian problem [57,66]. Since all commuting Pauli Hamiltonians are noncontextual, we will generalize this by proving that the noncontextual Hamiltonian problem is NP-complete.

### C. Noncontextual Hamiltonians

For $\mathcal{S}$ the set of Pauli operators in the Hamiltonian, let $\overline{\mathcal{S}}$ denote the *closure under inference* of $\mathcal{S}$ [17,47]. This means that $\overline{\mathcal{S}}$ is the smallest set containing $\mathcal{S}$ as a subset, such that for every commuting pair $A, B \in \overline{\mathcal{S}}$, $AB \in \overline{\mathcal{S}}$ as well. Since we could in principle measure $A$ and $B$ simultaneously together with their product $AB$, from any value assignment to $A, B$ we may *infer* the assignment to $AB$: hence closure under inference. We say that a value assignment is *consistent* if it respects all such inference relations. A value assignment to the observables in $\mathcal{S}$ (an *ontic state*) extends to an assignment to $\overline{\mathcal{S}}$, and noncontextuality is defined by situations in which consistent assignments to $\overline{\mathcal{S}}$ exist [17].

We can determine whether $\mathcal{S}$ is contextual by the following criterion. Let $\mathcal{Z} \subseteq \mathcal{S}$ be the set of operators in $\mathcal{S}$ that commute with all operators in $\mathcal{S}$ (we say that elements of $\mathcal{Z}$ commute universally in $\mathcal{S}$). Let $\mathcal{T} \equiv \mathcal{S} \setminus \mathcal{Z}$; then $\mathcal{S}$ *is noncontextual if and only if commutation is an equivalence relation on* $\mathcal{T}$ [17]. In other words, if $\mathcal{S}$ is noncontextual, then $\mathcal{T}$ is a union of $N$ disjoint "cliques" $C_1, C_2, \ldots, C_N$, such that operators in different cliques anticommute, while operators in the same clique commute.

Let $C_i \equiv \{C_{ij} \mid j = 1, 2, \ldots, |C_i|\}$. A noncontextual Hamiltonian then has the form

$$H = \sum_{i=1}^{N} \left( \sum_{j=1}^{|C_i|} h_{ij} C_{ij} \right) + \sum_{B \in \mathcal{Z}} h_B B, \tag{2}$$

where $h_{ij}$ and $h_B$ are real coefficients.

We may rewrite this Hamiltonian in a useful way by defining $A_{ij} \equiv C_{ij} C_{i1}$. $A_{ij}$ is itself a Pauli operator (up to a sign) and commutes universally in $\mathcal{S}$, since any operator in $C_i$ or in $\mathcal{Z}$ commutes with both $C_{ij}$ and $C_{i1}$, and any operator in one of the other cliques anticommutes with both $C_{ij}$ and $C_{i1}$. Thus, we may rewrite (2) as

$$H = \sum_{i=1}^{N} \left( \sum_{j=1}^{|C_i|} h_{ij} A_{ij} \right) C_{i1} + \sum_{B \in \mathcal{Z}} h_B B. \tag{3}$$

Note that in (3), the only operators appearing on the right-hand side that do not commute universally in $\mathcal{S}$ are the $C_{i1}$: $C_{i1}$ and $C_{i'1}$ anticommute for $i \neq i'$. See Appendix B for further discussion of noncontextual sets.

## II. QUASIQUANTIZED MODEL FOR A NONCONTEXTUAL HAMILTONIAN

In this section we describe our main tool, a quasiquantized (or *epistricted*) model for noncontextual Hamiltonians. A quasiquantized model is composed of a set of phase space points, called the *ontic states* of the system, and a set of probability distributions over the ontic states, called the *epistemic states* [49,50].

### A. Ontic states

If the set $\mathcal{S}$ is noncontextual, then consistent joint valuations for $\mathcal{S}$ exist. However, in general not *every* assignment of values to $\mathcal{S}$ is consistent. This motivates the following:

*Definition 1.* An *independent set* of Pauli operators contains no operator that can be written as a product of other commuting operators in the set.

In other words, a set is independent if it contains no operators whose values can be inferred from the values of other operators in the set, according to the notion of inference defined in Sec. I C. If a set of Pauli operators is independent, then *every* joint valuation for the set is consistent. Although the set $\mathcal{S}$ of Hamiltonian terms may be dependent, we can construct an independent set $\mathcal{R}$ such that $\overline{\mathcal{R}} = \overline{\mathcal{S}}$. Since $\mathcal{R}$ is independent, the joint valuations for $\mathcal{R}$ are $\{+1, -1\}^{|\mathcal{R}|}$: these label the ontic states.

To obtain $\mathcal{R}$, we first construct

$$G' \equiv \mathcal{Z} \cup \left( \bigcup_{i=1}^{N} \{A_{ij} \mid j = 2, 3, \ldots, |C_i|\} \right). \tag{4}$$

All operators in $G'$ commute universally in $\mathcal{S}$. $G'$ will in general be dependent, but we may obtain an independent set $G \equiv \{G_i \mid i = 1, 2, \ldots, |G|\}$ from $G'$ using the method described in [[67], Sec. 10.5.7] (summarized in Appendix C). Operators in $G$ commute universally in $\mathcal{S}$ because $G \subseteq \overline{G'}$.

The independent set $\mathcal{R}$ is then given by

$$\mathcal{R} \equiv \{C_{i1} \mid i = 1, 2, \ldots, N\} \cup G. \tag{5}$$

The $C_{i1}$ are independent because no pair of them commutes, and since any product of operators in $G$ commutes with all operators in $\mathcal{S}$, the $C_{i1}$ are also independent of $G$. Therefore, $\mathcal{R}$ is an independent set.

Each operator in $\mathcal{Z}$ and each $A_{ij}$ may be written as a product of operators in $G$ (see Appendix C). Therefore, since

$$\mathcal{T} = \bigcup_{i=1}^{N} \{A_{ij} C_{i1} \mid j = 1, 2, \ldots, N\}, \tag{6}$$

each operator in $\mathcal{T}$ is a product of one of the $C_{i1}$ and some set of operators in $G$. Thus, each operator in $\mathcal{S}$ is a product of commuting operators in $\mathcal{R}$, so $\overline{\mathcal{R}} = \overline{\mathcal{S}}$.

The set $G$ has size at most $n-1$, because a set of $n$ independent, commuting Pauli operators forms a complete, commuting set of observables for $n$ qubits. Therefore, if $G$ had size $n$ (or more), each set $G \cup \{C_{i1}\}$ would be a commuting set of size $n+1$ (or more), which could not be independent. Thus $\mathcal{R}$ has size at most $N + n - 1$. As pointed out in [[47], Sec. IV A], $N \leqslant 2m + 1$ for $m \equiv n - |G|$, so $|\mathcal{R}|$ is in fact at most

$2n+1$, which can be true only when $G$ (and thus $\mathcal{Z}$) is empty. This will be important in Secs. III and IV.

Note that although $\mathcal{R}$ is not in general a subset of $\mathcal{S}$, it is a subset of $\overline{\mathcal{S}}$. Furthermore, since there is a bijection from the set of ontic states for $\mathcal{R}$ to the set of ontic states for $\overline{\mathcal{S}}$, $|\mathcal{R}|$ is unique for each $\overline{\mathcal{S}}$. Thus closure under inference permits construction of the independent generating set $\mathcal{R}$.

### B. Epistemic states

Epistemic states are joint probability distributions over the ontic states, which complete our quasiquantized model. We write these joint probabilities as $P(c_1, \ldots, c_N, g_1, g_2, \ldots)$, where each $c_i$, $g_i$ is $\pm 1$ and denotes the value assigned to $C_{i1}$ or $G_i$, respectively.

First consider a commuting Hamiltonian. In this case, $\mathcal{R} = G$, and $\overline{\mathcal{S}} = \overline{\mathcal{R}}$ is the Abelian group generated by $\mathcal{R}$. The observables may be simultaneously measured, so there is a one-to-one mapping between the ontic states and the simultaneous eigenstates of $\mathcal{S}$. Thus the only constraint on the joint probabilities in this case is normalization.

Next consider the case where all observables pairwise anticommute:

*Lemma 1.* Let $\vec{A} = (A_1, A_2, \ldots, A_N)$ be an anticommuting set of Pauli operators. For any unit vector $\vec{a} \in \mathbb{R}^N$, the operator $\sum_{i=1}^N a_i A_i$ has eigenvalues $\pm 1$. From this it follows that for any state, $\sum_{i=1}^N \langle A_i \rangle^2 \leqslant 1$.

We prove Lemma 1 in Appendix A.

For a general noncontextual set $\mathcal{S}$, construct $\mathcal{R}$ as described in Sec. II A. The set $E = E(\mathcal{R})$ of epistemic states is then

$$E \equiv \left\{ (\vec{q}, \vec{r}) \in \{\pm 1\}^{|G|} \times \mathbb{R}^N \,|\, |\vec{r}| = 1 \right\}. \tag{7}$$

The pairs $(\vec{q}, \vec{r})$ define the joint probabilities as follows:

$$P_{(\vec{q},\vec{r})}(c_1, \ldots, c_N, g_1, g_2, \ldots) = \left( \prod_{j=1}^{|G|} \delta_{g_j, q_j} \right) \prod_{i=1}^{N} \frac{1}{2} |c_i + r_i|. \tag{8}$$

We refer to both the joint probabilities and the vector pairs $(\vec{q}, \vec{r})$ as epistemic states: they contain equivalent information.

In terms of $(\vec{q}, \vec{r})$, the expectation values for $\mathcal{R}$ are given by

$$\begin{aligned} \langle G_j \rangle_{(\vec{q},\vec{r})} &= q_j, \\ \langle C_{i1} \rangle_{(\vec{q},\vec{r})} &= r_i. \end{aligned} \tag{9}$$

*Theorem 1.* For epistemic states $(\vec{q}, \vec{r})$ as defined in (7), the joint probability distribution (8) is equivalent to the set of expectation values (9).

We prove Theorem 1 in Appendix A.

The model (7) is epistricted in the following sense: as in [49], a state is represented by joint knowledge of a set of commuting observables. For a given $(\vec{q}, \vec{r})$, this set is $G$ together with the observable

$$\mathcal{A}(\vec{r}) \equiv \sum_{i=1}^{N} r_i C_{i1} \tag{10}$$

(which has eigenvalues $\pm 1$, by Lemma 1). Note that since the $C_{i1}$ have expectation values $r_i$ as in (9), $\mathcal{A}(\vec{r})$ has expectation value 1, since $\vec{r}$ is a unit vector. No probability distributions are allowed that represent more knowledge of the state than simultaneous values of $G$ and $\mathcal{A}(\vec{r})$. Note that our model describes only pure states (as do the models in [49]).

From the expectation values (9) for $\mathcal{R}$, we can obtain expectation values for $\mathcal{S}$ as follows. For $B \in \mathcal{Z}$, let $\mathcal{J}_B$ be the set of indices such that $B = \prod_{j \in \mathcal{J}_B} G_j$; then

$$\langle B \rangle_{(\vec{q},\vec{r})} = \left\langle \prod_{j \in \mathcal{J}_B} G_j \right\rangle = \prod_{j \in \mathcal{J}_B} q_j, \tag{11}$$

where the second equality follows because $G_j \mapsto q_j = \pm 1$ for all $j$ (in other words, the state is a common eigenstate of the $G_j$ and of $B$). Similarly, for $C_{i1} B \in \mathcal{T}$,

$$\langle C_{i1} B \rangle_{(\vec{q},\vec{r})} = r_i \prod_{j \in \mathcal{J}_B} q_j. \tag{12}$$

*Theorem 2.* The epistemic states (7) give sets of expectation values that correspond to valid quantum states, and the set of quantum states described by the epistemic states includes an eigenbasis of any Hamiltonian whose Pauli terms are $\mathcal{S}$.

The proof of Theorem 2 is given in Appendix A. Note that for any $(\vec{q}, \vec{r})$, the expectation values (9) are produced by a simultaneous eigenstate of $G \cup \{\mathcal{A}(\vec{r})\}$. For the second claim in Theorem 2, we show that there exists an eigenbasis for the Hamiltonian composed of common eigenstates of $G$ and $\mathcal{A}(\vec{r})$ for some $\vec{r}$. That $\mathcal{A}(\vec{r})$ may be included is implied by:

*Lemma 2.* For $|\psi\rangle$ an eigenstate of the full Hamiltonian (3), the expectation values of the $C_{i1}$ satisfy $\sum_{i=1}^N \langle C_{i1} \rangle^2 = 1$.

The proof of Lemma 2 is given in Appendix A. In other words, the $\langle C_{i1} \rangle$ saturate the bound given in Lemma 1, for any energy eigenstate. This means that every energy eigenvalue can be reproduced via the expectation values (11) and (12) for some setting of $(\vec{q}, \vec{r})$.

We show in Appendix D how for *any* quantum state we may construct a joint probability distribution that reproduces the expectation values for $\overline{\mathcal{S}}$; however, to simulate noncontextual VQE it is only necessary to reproduce probabilities corresponding to eigenstates.

## III. CLASSICAL SIMULATION OF A NONCONTEXTUAL HAMILTONIAN

### A. Classical objective function

Given the model described in Sec. II, we now define a classical variational algorithm to simulate a noncontextual Hamiltonian. In (3), each $A_{ij}$ and each $B$ is a product of operators in $G$, i.e., is an element of $\overline{G}$. Therefore, we may replace $A_{ij}$ by $B$ and sum over all of $\overline{G}$, obtaining

$$H = \sum_{B \in \overline{G}} \left( h_B B + \sum_{i=1}^{N} h_{B,i} B C_{i1} \right), \tag{13}$$

where the $h_{B,i}$ and $h_B$ are just relabelings of the coefficients in (2). Thus we can use (11) and (12) to write:

$$\langle H \rangle_{(\vec{q},\vec{r})} = \sum_{B \in \overline{G}} \left( h_B + \sum_{i=1}^{N} h_{B,i} r_i \right) \prod_{j \in \mathcal{J}_B} q_j. \tag{14}$$

We may now treat (14) as a classical objective function.

This classical optimization problem will in general be hard. Although a convex special case of (14) is obtained when

TABLE I. Contextual VQE experiments, as approximated by noncontextual and diagonal Hamiltonians. $n$ is the number of qubits. $|\mathcal{S}_{\text{full}}|$ is the number of terms in the full Hamiltonian, $|\mathcal{S}_{\text{noncon}}|$ is the number of terms in the noncontextual sub-Hamiltonian, and $|\mathcal{R}|$ is the number of parameters in an epistemic state (which is upper bounded by $2n+1$ for $n$ qubits). $\epsilon_{\text{noncon}}$ is the error in the noncontextual approximation, $\epsilon_{\text{diag}}$ is the error obtained by only keeping the diagonal terms in the Hamiltonian, and $\epsilon_{\text{expt}}$ is the error in the VQE experiment. Errors are in units of chemical accuracy, 0.0016 Ha. Experimental errors preceded by $\sim$ were estimated from figures.

| Citation | System | $n$ | $\lvert\mathcal{S}_{\text{full}}\rvert$ | $\lvert\mathcal{S}_{\text{noncon}}\rvert$ | $\lvert\mathcal{R}\rvert$ | $\epsilon_{\text{noncon}}$ | $\epsilon_{\text{diag}}$ | $\epsilon_{\text{expt}}$ | Expt. outperforms noncontextual? |
|---|---|---|---|---|---|---|---|---|---|
| Peruzzo *et al.*, 2014 [2] | $HeH^+$ | 2 | 9 | 5 | 3 | 0.21 | 4.1 | 4.1 | No |
| Hempel *et al.*, 2018 [11] | LiH | 3 | 13 | 9 | 4 | 0.56 | 0.56 | $\sim$80 | No |
| Kandala *et al.*, 2017 [10] | LiH | 4 | 99 | 23 | 5 | 4.2 | 9.3 | $\sim$30 | No |
| Kandala *et al.*, 2017 [10] | $BeH_2$ | 6 | 164 | 42 | 7 | 156 | 266 | $\sim$90 | Yes |

we fix some set of values for the $q_j$ (or when $G$ is empty), in general the function is nonconvex. It is also in general frustrated, even if all terms commute, since in that case (14) becomes a linear combination of products of the $q_j$. Thus, as discussed further in Sec. III B, we should not expect worst cases of this optimization to be tractable, but they remain at worst classically hard.

### B. The noncontextual Hamiltonian problem

The statement of the noncontextual Hamiltonian problem is as follows: The inputs are a Hamiltonian $H$ of the form (13) with poly($n$) terms, together with a "promise" that the lowest eigenvalue of $H$ is either greater than $b$ or less than $a$ for some $a < b \in \mathbb{R}$ such that $b - a > 1/\text{poly}(n)$. The goal is to return YES if the lowest eigenvalue is less than $a$.

A problem is in NP if for every YES instance, there exists a proof (or *witness*) that is classically verifiable in polynomial time. In our case, the witness for a YES instance is an epistemic state $(\vec{q}, \vec{r}) \in E$ (a vector with dimension at most $2n+1$, as discussed in Sec. II A) satisfying $\langle H \rangle_{(\vec{q},\vec{r})} < a$. By showing that such a witness can be efficiently verified, we prove the following in Appendix A:

*Theorem 3.* The noncontextual Hamiltonian problem is in NP.

The diagonal local Hamiltonian problem is NP-complete, as follows from [53]; see also [63]. This remains true even for 2-local diagonal Hamiltonians [55], so since any 2-local diagonal Hamiltonian is noncontextual with poly($n$) Pauli terms, the noncontextual Hamiltonian problem is NP-complete as well.

## IV. APPROXIMATION OF GENERAL HAMILTONIANS BY NONCONTEXTUAL HAMILTONIANS

Can we use our classical simulation technique as an approximation method for contextual Hamiltonians? Since the model described in Sec. II depends structurally on the Hamiltonian being noncontextual, we cannot apply it to a general Hamiltonian directly. However, given a general Hamiltonian we can find a noncontextual subset of the terms, and take their (collective) ground state energy as an approximation of the true value.

Finding the largest noncontextual subset of terms is a generalization of the disjoint cliques problem [68], which is NP-complete. However, to date VQE experiments have largely focused on Hamiltonians in which the total weight of the terms (in $l_1$ norm) is dominated by the diagonal terms. Therefore, as a heuristic we select terms from the full Hamiltonian greedily by coefficient magnitude while the set remains noncontextual, thus obtaining the diagonal terms together with some additional set of relatively low weight. Given this noncontextual set, we construct $\mathcal{R}$ and minimize the resulting objective function (14) by brute-force search, since for few qubits $\mathcal{R}$ is small (see Sec. II A and Table I). For larger examples a classical optimization technique should replace this brute-force search.

We applied this heuristic to contextual Hamiltonians that have been simulated in VQE experiments to date. The results are given in Table I. The best noncontextual sub-Hamiltonians we found, for each full Hamiltonian in Table I, are listed in Appendix E.

For small Hamiltonians, our noncontextual approximation reached chemical accuracy. In all but one case the noncontextual approximation outperformed the approximation obtained by keeping only the diagonal terms. This is a natural point of comparison, since diagonal Hamiltonians constitute another common notion of classicality, and any set of diagonal Pauli operators is noncontextual.

Also, in all cases except for the $BeH_2$ simulation in [10], the noncontextual approximation reached better accuracy than the corresponding experiment. We wish to be clear that this is not a criticism of these experiments, which were intended as demonstrations of methodology rather than as precise estimations. However, what our noncontextual approximations show is that these experiments have not achieved sufficient accuracy to resolve intrinsically quantum behavior, i.e., the full-configuration correction to the noncontextual ground state energy.

Finally, the fact that the $BeH_2$ experiment does outperform our approximation indicates that, as we would expect, more terms in the Hamiltonian means more room for contextuality, and hence worse noncontextual approximations. Thus, we may hope that future experiments simulating larger Hamiltonians will reliably exceed this minimum standard for quantum behavior. On the other hand, better heuristics for identifying the noncontextual set may improve the noncontextual approximation.

## V. DISCUSSION

In the quantum approximate optimization algorithm (QAOA) [69], the Hamiltonian is diagonal (and thus noncontextual), because it encodes a classical problem, so our

method simply recovers the diagonal entries. We have shown that the noncontextual Hamiltonian problem is in NP. Thus, the potential for quantum advantage in noncontextual VQE reduces to the same question that motivates QAOA: can we use cleverly chosen and/or physically motivated ansatze to generate otherwise hard-to-reach joint probability distributions, and thus efficiently converge to solutions of classically hard problems?

We demonstrated in Sec. IV that our model is applicable as an approximation method for contextual Hamiltonians, such as general electronic-structure Hamiltonians. This technique provides a more stringent test for nonclassicality than that given Table I in [17], by demonstrating that some experiments with contextual Hamiltonians do not achieve sufficient accuracy to tell their results apart from those due to a noncontextual approximation. Again, we wish to stress that this is not a criticism of these experiments, which have played seminal roles in the development of quantum simulation techniques, but only a means by which we may try to identify intrinsically quantum behavior.

Finally, in addition to serving as a benchmark for quantum experiments, our simulation technique may be useful as a new approximation method in its own right. Also, it may be possible to extend our criterion for noncontextuality to other Hamiltonian decompositions besides Pauli decomposition, thus improving the capacity of our simulation algorithm. We leave the full exploration of these possibilities for future work.

## ACKNOWLEDGMENTS

W.M.K. acknowledges support from the National Science Foundation, Grant No. DGE-1842474. P.J.L. acknowledges support from the National Science Foundation, Grant No. PHY-1720395, and from Google Inc. This work was supported by the National Science Foundation STAQ project (PHY-1818914).

## APPENDIX A: PROOFS

*Lemma 1.* Let $\vec{A} = (A_1, A_2, \ldots, A_N)$ be an anticommuting set of Pauli operators. For any unit vector $\vec{a} \in \mathbb{R}^N$, the operator $\vec{a}\cdot\vec{A} \equiv \sum_{i=1}^N a_i A_i$ has eigenvalues $\pm 1$. From this it follows that for any state, $\sum_{i=1}^N \langle A_i\rangle^2 \leqslant 1$.

*Proof.* This theorem consists of two claims:

For the first claim (that the operator $\vec{a}\cdot\vec{A}$ has eigenvalues $\pm 1$), note that

$$\begin{aligned}(\vec{a}\cdot\vec{A})^2 &= \sum_i a_i^2 A_i^2 + \sum_{i>j} a_i a_j \{A_i, A_j\} \\ &= \sum_i a_i^2 \mathbb{1} = |\vec{a}|^2 \mathbb{1} = \mathbb{1},\end{aligned} \tag{A1}$$

where the second equality follows because the Pauli operators $A_i$ are self-inverse, and by assumption $\{A_i, A_j\} = 0$ for $i > j$. Thus, $\vec{a}\cdot\vec{A}$ is self-inverse, so its eigenvalues are $\pm 1$. This completes the proof of the first claim.

For the second claim (that $\sum_{i=1}^N \langle A_i\rangle^2 \leqslant 1$), note that for any state represented as a density operator $\rho$, we may write

$$\rho = \tfrac{1}{2^n}(\mathbb{1} + \vec{b}\cdot\vec{A} + \cdots), \tag{A2}$$

for some $\vec{b} \in \mathbb{R}^N$, where the ellipsis indicates the presence in general of additional Pauli terms. The expectation value of any $A_i$ is then

$$\langle A_i\rangle = \mathrm{Tr}(A_i\rho) = b_i, \tag{A3}$$

since the Pauli operators are Hilbert-Schmidt orthogonal and self-inverse.

Assume that $|\vec{b}| > 0$ [if $|\vec{b}| = 0$, then the theorem holds, by (A3)]. Let $\vec{a} \equiv \frac{\vec{b}}{|\vec{b}|}$, so that $|\vec{a}| = 1$. By the first claim in this lemma, the observable $\vec{a}\cdot\vec{A}$ has eigenvalues $\pm 1$. Therefore, $|\langle\vec{a}\cdot\vec{A}\rangle| \leqslant 1$, so we have

$$\begin{aligned}1 &\geqslant |\langle\vec{a}\cdot\vec{A}\rangle| = |\mathrm{Tr}(\vec{a}\cdot\vec{A}\rho)| \\ &= \left|\frac{1}{2^n}\mathrm{Tr}\big(\vec{a}\cdot\vec{A}(\vec{b}\cdot\vec{A})\big)\right| = |\vec{a}\cdot\vec{b}|,\end{aligned} \tag{A4}$$

where (as above) the steps in the second line follow because the Pauli operators are Hilbert-Schmidt orthogonal and self-inverse. Therefore,

$$1 \geqslant |\vec{a}\cdot\vec{b}|^2 = \frac{|\vec{b}|^4}{|\vec{b}|^2} = |\vec{b}|^2 = \sum_{i=1}^N b_i^2 = \sum_{i=1}^N \langle A_i\rangle^2 \tag{A5}$$

(since by assumption $|\vec{b}| > 0$). This completes the proof of the second claim. ■

*Theorem 1.* For epistemic states $(\vec{q}, \vec{r})$ as defined in (7), the joint probability distribution (8) is equivalent to the set of expectation values (9).

*Proof.* (1) We first prove the reverse implication: assume that (9) holds. We reproduce (9) here for convenience:

$$\begin{aligned}\langle G_j\rangle_{(\vec{q},\vec{r})} &= q_j, \\ \langle C_{i1}\rangle_{(\vec{q},\vec{r})} &= r_i.\end{aligned} \tag{A6}$$

That these lead to the joint probabilities given by (8) essentially follows from the discussion following (9), but we will fill in the details.

Since $\langle G_j\rangle = q_j = \pm 1$ for each $j$ [as in (7)], the $G_j$ have definite values. The values of the $G_j$ in the ontic state $(c_1, \ldots, c_N, g_1, g_2, \ldots)$ are the $g_j$, so $P(c_1, \ldots, c_N, g_1, g_2, \ldots)$ can be nonzero only when $q_j = g_j$ for each $j$. If this holds for each $j$, then $P(c_1, \ldots, c_N, g_1, g_2, \ldots)$ is just the product over $i = 1, 2, \ldots, N$ of the probabilities of obtaining outcome $C_{i1} \mapsto c_i$ given the expectation value $\langle C_{i1}\rangle = r_i$: these probabilities are given by

$$\tfrac{1}{2}|c_i + r_i| \tag{A7}$$

for each $i$. Taking the joint probability to be the product of these works only because the $C_{i1}$ do not commute, and thus cannot be correlated.

The condition due to the $G_j$ thus gives a factor of

$$\prod_{j=1}^{|G|} \delta_{g_j, q_j}, \tag{A8}$$

and the condition due to the $C_{i1}$ gives a factor of

$$\prod_{i=1}^{N} \frac{1}{2}|c_i + r_i|; \tag{A9}$$

(8) is simply the product of these.

Now assume that (8) holds. We reproduce (8) here for convenience:

$$P(c_1, \ldots, c_N, g_1, g_2, \ldots) = \left(\prod_{j=1}^{|G|} \delta_{g_j,q_j}\right)\left(\prod_{i=1}^{N} \frac{1}{2}|c_i + r_i|\right). \tag{A10}$$

The probabilities for the outcomes of each $G_j$ and each $C_{i1}$ should be obtained as marginals of (A10). If $p_{G_j}$ denotes the probability of obtaining the outcome $G_j \mapsto +1$, then

$$\begin{aligned} p_{G_j} &= \sum_{\substack{c_i, g_k = \pm 1, \\ \forall k \neq j}} P(c_1, \ldots, c_N, g_1, g_2, \ldots)\Big|_{g_j=1} \\ &= \sum_{\substack{c_i, g_k = \pm 1, \\ \forall k \neq j}} \left(\prod_{l=1}^{|G|} \delta_{g_l,q_l}\right)\Big|_{g_j=1} \left(\prod_{m=1}^{N} \frac{1}{2}|c_m + r_m|\right) \\ &= \delta_{q_j,1} \sum_{c_i=\pm 1} \left(\prod_{m=1}^{N} \frac{1}{2}|c_m + r_m|\right) \\ &= \frac{1}{2}(q_j + 1)\left(\prod_{m=1}^{N} \frac{1}{2}(|r_m + 1| + |r_m - 1|)\right) \\ &= \frac{1}{2}(q_j + 1), \end{aligned} \tag{A11}$$

where the fourth equality follows because for $r_m \in [-1, 1]$,

$$|r_m + 1| + |r_m - 1| = 2. \tag{A12}$$

Similarly,

$$\begin{aligned} p_{C_{i1}} &= \sum_{\substack{c_k, g_j = \pm 1, \\ \forall k \neq i}} P(c_1, \ldots, c_N, g_1, g_2, \ldots)\Big|_{c_i=1} \\ &= \sum_{\substack{c_k, g_j = \pm 1, \\ \forall k \neq i}} \left(\prod_{l=1}^{|G|} \delta_{g_l,q_l}\right)\left(\prod_{m=1}^{N} \frac{1}{2}|c_m + r_m|\right)\Big|_{c_i=1} \\ &= \sum_{\substack{c_k = \pm 1, \\ \forall k \neq i}} \left(\prod_{m=1}^{N} \frac{1}{2}|c_m + r_m|\right)\Big|_{c_i=1} \\ &= \frac{1}{2}|r_i + 1| = \frac{1}{2}(r_i + 1), \end{aligned} \tag{A13}$$

again using (A12). From (A11) and (A13) we can obtain the expectation values for $G_j$ and $C_{i1}$:

$$\begin{aligned} \langle G_j \rangle_{(\vec{q},\vec{r})} &= 2p_{G_j} - 1 = q_j, \\ \langle C_{i1} \rangle_{(\vec{q},\vec{r})} &= 2p_{C_{i1}} - 1 = r_i, \end{aligned} \tag{A14}$$

which is (9) [and (A6), above], as desired. ■

*Lemma 2.* For $|\psi\rangle$ an eigenstate of the full Hamiltonian (3), the expectation values of the $C_{i1}$ satisfy

$$\sum_{i=1}^{N} \langle C_{i1} \rangle^2 = 1. \tag{A15}$$

*Proof.* As discussed in the main text, we may simultaneously diagonalize the $G_j$. The resulting form for the full Hamiltonian $H$ will be block diagonal, with each block corresponding to a common eigenspace of the $G_j$, since the $C_{i1}$ commute with the $G_j$ and therefore do not mix their common eigenspaces. Thus, the eigenstates of the full Hamiltonian are common eigenstates of $G$.

Given this, there exists an eigenbasis of $H$ in which each eigenvector $|\psi\rangle$ is also an eigenstate of any $B \in \overline{G}$, a product of some subset of $G$. Therefore,

$$BC_{i1}|\psi\rangle = C_{i1}B|\psi\rangle = \lambda_B C_{i1}|\psi\rangle, \tag{A16}$$

where $\lambda_B = \pm 1$ is the eigenvalue of $|\psi\rangle$ for the operator $B$. Thus, we may rewrite the full Hamiltonian (3), acting on $|\psi\rangle$, as

$$H|\psi\rangle = \left(h'_0 + \sum_{i=1}^{N} h'_i C_{i1}\right)|\psi\rangle, \tag{A17}$$

for coefficients $h'_i$ defined by

$$\begin{aligned} h'_0 &\equiv \sum_{B \in \mathcal{Z}} h_B \lambda_B, \\ h'_i &\equiv \sum_{B \in \overline{G}} h_{B,i} \lambda_B \quad \text{for } i > 0, \end{aligned} \tag{A18}$$

where $h_{B,i}$ is the coefficient of $BC_{i1}$ in $H$. Since $|\psi\rangle$ is an eigenstate of $H$, by (A17) it must also be an eigenstate of

$$\sum_{i=1}^{N} h'_i C_{i1} = h \sum_{i=1}^{N} \tilde{h}_i C_{i1}, \tag{A19}$$

the nonidentity terms in (A17), for $h \equiv \sqrt{\sum_{i=1}^{N} (h'_i)^2}$, so that the $\tilde{h}_i \equiv h'_i/h$ satisfy

$$\sum_{i=1}^{N} \tilde{h}_i^2 = 1. \tag{A20}$$

Thus by Lemma 1, the operator $\sum_{i=1}^{N} \tilde{h}_i C_{i1}$ has eigenvalues $\pm 1$. Since $|\psi\rangle$ is an eigenstate of the operator given in (A19), it is an eigenstate of $\sum_{i=1}^{N} \tilde{h}_i C_{i1}$:

$$\left(\sum_{i=1}^{N} \tilde{h}_i C_{i1}\right)|\psi\rangle = \pm|\psi\rangle. \tag{A21}$$

Therefore, if $\langle\cdot\rangle$ denotes expectation value with respect to $|\psi\rangle$,

$$\langle\psi|\left(\sum_{i=1}^{N} \tilde{h}_i C_{i1}\right)|\psi\rangle = \sum_{i=1}^{N} \tilde{h}_i \langle C_{i1} \rangle = \pm 1. \tag{A22}$$

We know by 1 that

$$\sum_{i=1}^{N} \langle C_{i1} \rangle^2 \leqslant 1, \tag{A23}$$

TABLE II. $HeH^+$ Hamiltonian in Peruzzo *et al.*, 2014 [2]. Terms are displayed in the format {Pauli operator: coefficient,...}, i.e., as a Python dict mapping Pauli operators that appear in the Hamiltonian to their coefficients. Coefficients are given in Hartree.

**Diagonal terms:** {II: −1.46658, IZ: −0.39863, ZI: −0.39863, ZZ: 0.089735}

**Additional terms in noncontextual sub-Hamiltonian:** {XX: 0.099524}

**Additional terms in full Hamiltonian:** {IX: −0.087145, XI: −0.087145, XZ: 0.087145, ZX: 0.087145}

**Minimal energy parameter setting for $\mathcal{R}$, in noncontextual sub-Hamiltonian:**

$q_1 = \langle ZZ\rangle \mapsto +1, \quad r_1 = \langle XX\rangle \mapsto -0.1238712791070418, \quad r_2 = \langle IZ\rangle \mapsto 0.9922982949760547$

and by construction $\sum_{i=1}^{N} \tilde{h}_i^2 = 1$. Thus the only way (A22) can be satisfied is if

$$\langle C_{i1}\rangle = s\tilde{h}_i \tag{A24}$$

for all $i = 1, 2, \ldots, N$, for fixed $s = \pm 1$. Therefore,

$$\sum_{i=1}^{N} \langle C_{i1}\rangle^2 = \sum_{i=1}^{N} \tilde{h}_i^2 = 1. \tag{A25}$$

Since this holds for any $|\psi\rangle$ in the eigenbasis of $H$, it holds for any eigenstate of $H$. ■

*Theorem 2.* The epistemic states (7) give sets of expectation values that correspond to valid quantum states, and the set of quantum states described by the epistemic states includes an eigenbasis of any Hamiltonian whose Pauli terms are $\mathcal{S}$.

*Proof.* The theorem consists of two claims. For the first, note that for any epistemic state $(\vec{q}, \vec{r})$, there exists a simultaneous eigenstate of the operators $G_j$, with eigenvalue $q_j = \pm 1$ for each $G_j$, and of $\mathcal{A}(\vec{r}) \equiv \sum_{i=1}^{N} r_i C_{i1}$, with eigenvalue 1. For this state, the expectation value of each $G_j$ is $q_j$, and the expectation value of each $C_{i1}$ is $r_i$, as noted in Sec. II B. Thus every epistemic state $(\vec{q}, \vec{r})$ corresponds to a valid quantum state, proving the first claim.

For the second claim, consider first the universally commuting operators: $\mathcal{Z}$, which are generated by $G$. Since these may in principle be simultaneously diagonalized, the common eigenstates of $G$ (which are the common eigenstates for all of $\mathcal{Z}$) are a complete set of eigenstates for any Hamiltonian whose terms are a linear combination of $\mathcal{Z}$.

Now suppose we add to such a Hamiltonian a linear combination of the terms in $\mathcal{T}$, each of which is a product of operators in $G$ with one of the $C_{i1}$. As in the previous paragraph, we may simultaneously diagonalize $G$: the resulting form for the full Hamiltonian will be block diagonal, with each block corresponding to a common eigenspace of the $G_j$. Within each block, the Hamiltonian will take the form of some linear combination of the $C_{i1}$, obtained by replacing the $G_j$ by their eigenvalues for the current block, as in (A17) and (A18). Since the $C_{i1}$ commute with $G$, we may diagonalize the linear combination of the $C_{i1}$ within each block (i.e., the $C_{i1}$ do not mix the common eigenspaces of $G$). Thus, it is still the case that we may take the eigenstates of the full Hamiltonian to be common eigenstates of $G$. This justifies the condition on $\vec{q}$ (the first set of values in our epistemic states) given in (7), namely, $q_j = \pm 1$, since these correspond to the expectation values of the $G_j \in G$.

Regarding the condition on $\vec{r}$ (the second set of values in our epistemic state) in (7), we draw on Lemma 2. In (7), the condition on $\vec{r}$ is that it be a unit vector, so since its components $r_i$ give the expectation values for the $C_{i1}$ as in (9), Lemma 2 proves that the set of expectation values thus described includes all eigenvalues of the Hamiltonian. ■

*Theorem 3.* The noncontextual Hamiltonian problem is in NP.

*Proof.* We demonstrated in Sec. III A that by varying over all epistemic states, the expected energy given in (14) varies over all eigenvalues of the Hamiltonian. Therefore, if the Hamiltonian possesses an eigenvalue $\lambda < a$, there is some epistemic state $(\vec{q}, \vec{r})$ such that $\langle H\rangle_{(\vec{q},\vec{r})} = \lambda$. This epistemic state serves as a proof of the YES instance, as long as given the Hamiltonian as in (13) we can classically evaluate the objective function in (14) in polynomial time.

To show this, we use the fact that by assumption $|\mathcal{S}|$ (the number of terms in the Hamiltonian) is polynomial in $n$. Thus, the total number of nonzero coefficients $h_B$ and $h_{B,i}$ is poly($n$); these are given in the statement of the problem instance. Which coefficients are nonzero also determines the terms in the sum in (14) (and upper bounds their number). The only remaining components of (14) to be evaluated are the sets of indices $\mathcal{J}_B$: we have one such set for each term, defined to satisfy $B = \prod_{G\in\mathcal{J}_B} G$. The $\mathcal{J}_B$ are obtained directly from the standard method used to construct $G$ (described in Appendix C), the entirety of which requires poly($n$) classical operations given that $|\mathcal{S}| = \text{poly}(n)$. Thus, given a Hamiltonian of the form (13) and a witness in the form of an epistemic state $(\vec{q}, \vec{r})$, we can use (14) to verify that $\langle H\rangle_{(\vec{q},\vec{r})} < a$, in poly($n$) classical operations. ■

## APPENDIX B: NONCONTEXTUAL SETS OF PAULI OPERATORS

As noted in Sec. I C, the key concept in reasoning about noncontextuality is inference among value assignments to Pauli operators. This stems from the following property that we demand of a noncontextual ontological model for a set of Pauli observables: in any set of measurements we can perform simultaneously, the values we obtain from the noncontextual model must agree with those required by the full formalism of quantum mechanics. In other words, if $A$ and $B$ are commuting Pauli operators, then by measuring them simultaneously we can predict with certainty the result of measuring their product $AB$; therefore, in an assignment of values to $\{A, B, AB\}$, the value assigned to $AB$ must be the product of the values assigned to $A$ and $B$.

This motivates the closure under inference $\overline{\mathcal{S}}$ of a set $\mathcal{S}$ of Pauli operators, as defined in Sec. I C: it is the set of Pauli operators whose values are determined by a value assignment to $\mathcal{S}$ [17]. The Jordan product $\frac{1}{2}\{\cdot,\cdot\}$ provides another definition

TABLE III. LiH Hamiltonian in Hempel *et al.*, 2018 [11]. Terms are displayed in the format {Pauli operator: coefficient,...}, i.e., as a Python dict mapping Pauli operators that appear in the Hamiltonian to their coefficients. Coefficients are given in Hartree.

**Diagonal terms:** {III: −6.823060333, ZII: −0.1110098029, IZI: −0.5370907285, IIZ: −0.3127149146, ZZI: 0.383637914, ZIZ: 0.2581256772, IZZ: 0.2523178271}

**Additional terms in noncontextual sub-Hamiltonian:** {XXI: 0.06593809513, YYI: −0.06593809513}

**Additional terms in full Hamiltonian:** {XIX: 0.0121680127, YIY: 0.0121680127, IXX: 0.01764480014, IYY: 0.01764480014}

**Minimal energy parameter setting for $\mathcal{R}$, in noncontextual sub-Hamiltonian:**

$q_1 = \langle IIZ\rangle \mapsto +1, \quad q_2 = \langle ZZI\rangle \mapsto -1,$

$r_1 = \langle XXI\rangle \mapsto -0.000000129227, \quad r_2 = \langle ZII\rangle \mapsto -0.999999999999$

of $\overline{\mathcal{S}}$. Since a pair of Pauli operators $A, B$ either commute or anticommute,

$$\frac{1}{2}\{A, B\} = \begin{cases} AB \text{ if } [A, B] = 0, \\ 0 \text{ otherwise.} \end{cases} \tag{B1}$$

Thus, $\overline{\mathcal{S}}$ is the closure of $\mathcal{S}$ under the Jordan product.

A set of Pauli operators is noncontextual if ontic states for it exist that are self-consistent. We check consistency of an ontic state by making sure that in the ontic state it induces for the closure under inference, none of the inferences are violated: in other words, the value assigned each each product $AB$ of a commuting pair $A, B$ is the product of the values assigned to the pair.

Since by definition an ontic state (assignment of values) for $\mathcal{S}$ extends to an ontic state for $\overline{\mathcal{S}}$, and the ontic state for $\overline{\mathcal{S}}$ includes the ontic state for $\mathcal{S}$, any pair of noncontextual sets of Pauli operators that share the same closure under inference have equivalent ontic states that may derived from each other. Thus, given an arbitrary noncontextual set of Pauli operators, we may transform it into a noncontextual set with a standard form that shares the same closure under inference.

The first move we can make in doing this is, given a commuting pair $A, B$, so replace $B$ by $AB$. Since Pauli operators are self-inverse, $B$ is given by the product of $A$ and $AB$, so from a value assignment to $A$ and $AB$ we can infer the value assignment to $B$, and thus recover the original set. The second move we can make is to remove any $C$ that is the product of some commuting $A, B$ that are also in the set. In obtaining the independent set $\mathcal{R}$ that we used to define our ontic states in Sec. II A, we simply made a sequence of such moves.

The Jordan product is again useful in describing independent sets according to Definition 1. If $\mathcal{S}$ commutes, then this definition reduces to that in [[67], Sec. 10.5.1]. For the general noncontextual case, we would like to relate Definition 1 to the notion of independence that applies to sets of generators of Abelian groups, where it means that removing any of the generators reduces the size of the generated group. Since the Jordan product is commutative, Definition 1 nearly reproduces the usual notion of independence for an Abelian group taking the group operation to be the Jordan product, but the Jordan product is not associative. This distinction aside, $\overline{\mathcal{S}}$ is the closure of $\mathcal{S}$ under the Jordan product, and correspondingly, a subset of $\overline{\mathcal{S}}$ that is independent in our sense is also independent as a generating set under the Jordan product.

Thus, given an arbitrary noncontextual set $\mathcal{S}$, $\mathcal{R}$ as constructed in Sec. II A is an independent generating set for $\overline{\mathcal{S}}$ under the Jordan product. The construction we gave in the main text may therefore be summarized: a set of Pauli operators is noncontextual if and only if the independent generating sets for its closure under inference are composed of a set of universally commuting operators ($G$), and a set of pairwise anticommuting operators $\{C_{i1} \mid i = 1, 2, \ldots, N\}$.

## APPENDIX C: OBTAINING AN INDEPENDENT SET OF PAULI MEASUREMENTS FROM A COMMUTING SET

Given a set $G'$ of commuting Pauli operators on $n$ qubits, we wish to obtain an independent commuting set $G$ such that every operator in $G'$ is a product of operators in $G$. Since $G$ commutes, $G$ is independent if and only if no operator in $G$ is a product of other operators in $G$ (see the discussion following Definition 1). Finding $G$ given $G'$ is a standard procedure, and a method for performing it is given by Nielsen and Chuang in [[67], Sec. 10.5.7]. We summarize the method here in terms of the language used in this work.

The method to calculate $G$ from $G'$ is a multiplicative variant of Gaussian elimination. Let $G' = \{g'_1, g'_2, \ldots, g'_m\}$, and write each $g'_i$ as

$$g'_i = h_{i1} \otimes h_{i2} \otimes \cdots \otimes h_{in}, \tag{C1}$$

where each $h_{ij}$ is a single-qubit Pauli operator (including the identity). We may then express $G'$ in an array as

$$G' = \begin{matrix} h_{11} & h_{12} & h_{13} & \cdots & h_{1n}, \\ h_{21} & h_{22} & h_{23} & \cdots & h_{2n}, \\ h_{31} & h_{32} & h_{33} & \cdots & h_{3n}, \\ & & & \vdots & \\ h_{m1} & h_{m2} & h_{m3} & \cdots & h_{mn}, \end{matrix} \tag{C2}$$

where we have suppressed the tensor product symbols.

Let $h$ denote the matrix whose entries are $h_{ij}$, the single-qubit Pauli operators appearing in (C2), augmented by a vector $\vec{s}$ whose entries $s_i$ are the signs associated with each row in $h_i$ (initially these are all $+1$). We first describe a procedure that, given such a matrix $h$, transforms it to a matrix in which at most two entries in the first column are nonidentity:

(1) $h_{11}, h_{21}, \ldots, h_{m1}$ are the entries in the first column of $h$. If

$$h_{11} = h_{21} = \cdots = h_{m1} = I, \tag{C3}$$

TABLE IV. LiH Hamiltonian in Kandala *et al.*, 2017 [10]. Terms are displayed in the format {Pauli operator: coefficient,...}, i.e., as a Python dict mapping Pauli operators that appear in the Hamiltonian to their coefficients. Coefficients are given in Hartree.

**Diagonal terms:** {ZIII: −0.096022, IZII: 0.364746, IIZI: 0.096022, IIIZ: −0.364746, ZZII: −0.206128, ZIZI: −0.145438, ZIIZ: 0.110811, IZZI: 0.110811, IZIZ: −0.095216, IIZZ: −0.206128, ZZZI: −0.056040, ZZIZ: 0.063673, ZIZZ: 0.056040, IZZZ: −0.063673, ZZZZ: 0.080334}

**Additional terms in noncontextual sub-Hamiltonian:** {YYZI: 0.039155, XXZI: −0.039155, YYII: 0.02964, XXII: −0.02964, YYIZ: −0.02428, XXIZ: 0.02428, YYZZ: 0.002895, XXZZ: −0.002895}

**Additional terms in full Hamiltonian:** {XZII: −0.012585, XIII: 0.012585, IIXZ: 0.012585, XZIZ: 0.007265, XIIZ: −0.007265, XZZI: −0.011962, XIZI: 0.011962, XZZZ: −0.000247, XIZZ: 0.000247, IIZZ: 0.012585, IIXI: 0.012585, XZXZ: −0.002667, XZXI: −0.002667, XIXZ: 0.002667, XIXI: 0.002667, IZXZ: 0.007265, IZXI: 0.007265, IXII: 0.002792, IIXX: −0.029640, IIIX: 0.002792, XIXX: −0.008195, XIIX: −0.001271, XXXI: −0.008195, XXXX: 0.028926, XXIX: 0.007499, IXXI: −0.001271, IXXX: 0.007499, IXIX: 0.009327, IIYY: 0.029640, YYYY: 0.028926, ZXII: 0.002792, IIZX: −0.002792, ZIZX: −0.016781, ZIIX: 0.016781, ZXZI: −0.016781, IXZI: −0.016781, ZXZX: −0.009327, ZXIX: 0.009327, IXZX: −0.009327, ZIXZ: −0.011962, ZIXI: −0.011962, ZZXZ: 0.000247, ZZXI: 0.000247, ZIXX: 0.039155, ZZXX: −0.002895, ZZIX: −0.009769, IZXX: −0.024280, IZIX: −0.008025, ZIYY: −0.039155, ZZYY: 0.002895, IZYY: 0.024280, XZXX: 0.008195, XZIX: 0.001271, XZYY: −0.008195, XIYY: 0.008195, XZZX: −0.001271, XIZX: 0.001271, IZZX: 0.008025, IXZZ: −0.009769, IXIZ: 0.008025, XXXZ: −0.008195, IXXZ: −0.001271, YYXZ: 0.008195, YYXI: 0.008195, XXYY: −0.028926, IXYY: −0.007499, YYXX: −0.028926, YYIX: −0.007499, XXZX: −0.007499, YYZX: 0.007499, ZZZX: 0.009769, ZXXZ: −0.001271, ZXXI: −0.001271, ZXIZ: 0.008025, ZXXX: 0.007499, ZXYY: −0.007499, ZXZZ: −0.009769}

**Minimal energy parameter setting for $\mathcal{R}$, in noncontextual sub-Hamiltonian:**

$q_1 = \langle IIIZ\rangle \mapsto +1, \quad q_2 = \langle ZZII\rangle \mapsto +1, \quad q_3 = \langle IIZI\rangle \mapsto -1,$

$r_1 = \langle YYZI\rangle \mapsto -0.2192200361485217, \quad r_2 = \langle IZII\rangle \mapsto -0.9756754459096738$

then we are already done.

(2) If for some $k$, $h_{k1} = X$, then for each $i \neq k$, if $h_{i1} = X$, multiply row $i$ by row $k$. This corresponds to multiplying $g_i'$ by $g_k'$, or the following mapping on the entries in $h$:

$$\begin{aligned} h_{i1} &\mapsto h_{i1}h_{k1} = X^2 = I, \\ h_{ij} &\mapsto h_{ij}h_{kj}, \quad \forall j > 1. \end{aligned} \tag{C4}$$

Thus when we have completed this step for each $i \neq k$, there will be no $X$s in the first column except in row $k$.

(3) If for some $l$, $h_{l1} = Z$, then for each $i \neq l$, if $h_{i1} = Z$, multiply row $i$ by row $l$. As in step 2, this corresponds to multiplying $g_i'$ by $g_l'$, so when we have completed this step for each $i \neq l$, there will be no $Z$s in the first column except in row $l$.

(4) If there is both an $X$ and a $Z$ in the first column (in rows $k$ and $l$, respectively), then for each $i \neq k, l$, if $h_{i1} = Y$, multiply row $i$ by row $k$ and row $l$. This corresponds to multiplying $g_i'$ by $g_k'g_l'$, or the following mapping on the entries in $h$:

$$\begin{aligned} h_{i1} &\mapsto h_{i1}h_{k1}h_{l1} = YXZ = -iI, \\ h_{ij} &\mapsto h_{ij}h_{kj}h_{lj}, \quad \forall j > 1. \end{aligned} \tag{C5}$$

Thus when we have completed this step for each $i \neq l$, there will be no $Y$s in the first column at all (if both $X$ and $Z$ are present in the first column).

(5) If $X$ and $Z$ are not both present in the first column, and if for some $m$, $h_{m1} = Y$, then for each $i \neq m$, if $h_{i1} = Y$, multiply row $i$ by row $m$. As in step 2, this corresponds to multiplying $g_i'$ by $g_m'$, so when we have completed this step for each $i \neq m$, there will be no $Y$s in the first column except in row $m$.

Note that since we know that the full Pauli operators corresponding to the rows commute, the total phase obtained in multiplying any row by any other row must be $\pm 1$. We record this by multiplying $s_i$ by this sign, when the multiplication takes place in the $i$th row.

Call this procedure ReduceFirstCol($h'$). When we have completed the mapping, the transformed matrix $h$ will contain at most two rows in which the first entry is nonidentity (and the first entries in those two rows will be different).

Our full procedure to obtain $G$ is then as follows:

(1) Let $h$ be the full matrix (C2). $h'$ will be a submatrix of $h$ that is updated at each iteration; let $h'$ initially be equal to $h$.

(2) Perform the mapping

$$h' \mapsto \text{ReduceFirstCol}(h'). \tag{C6}$$

(3) For each of the (up to) two rows in $h'$ in which the first entry is nonidentity, put the corresponding Pauli operators in $G$. Then let the new $h'$ be the submatrix obtained from the current $h'$ by removing these rows and the first column.

(4) If any nonidentity entries remain in $h'$, return to step 2. If not, then $G$ is complete.

When this procedure is complete, at most two rows in $h$ will have a nonidentity first entry, at most two others will have a nonidentity second entry, at most two others will have a nonidentity third entry, and so forth. In other words, under some reordering of the rows, $h$ will have the form

$$\begin{array}{cccccccc} P_1 & h'_{12} & h'_{13} & h'_{14} & h'_{15} & \cdots & h'_{1n}, \\ [P_2 & h'_{22} & h'_{23} & h'_{24} & h'_{25} & \cdots & h'_{2n},] \\ I & P_3 & h'_{33} & h'_{34} & h'_{35} & \cdots & h'_{3n}, \\ [I & P_4 & h'_{43} & h'_{44} & h'_{45} & \cdots & h'_{4n},] \\ I & I & P_5 & h'_{54} & h'_{55} & \cdots & h'_{5n}, \\ [I & I & P_6 & h'_{64} & h'_{65} & \cdots & h'_{6n},] \\ & & & \vdots & & & \\ [I & I & I & I & I & \cdots & I], \end{array} \tag{C7}$$

where the bracketed rows may or may not appear, and in cases where they do appear $P_2 \neq P_1$, $P_4 \neq P_3$, $P_6 \neq P_5$, and so forth. Note that in general there may be a collection of rows at the bottom of the matrix that have been reduced entirely to the identity. $G$ will be given by the nonidentity rows in (C7).

To see that $G$ is independent, consider some particular nonidentity row $i$ (representing an operator $g_i \in G$). Row $i$ cannot be written as a product of rows below it, since if $h_{ij}$ is row $i$'s first nonidentity entry, $h_{i'j} = I$ for all $i' = i+2, i+3, \ldots, m$. The entry $h_{(i+1)j}$ (immediately below $h_{ij}$) might not be the identity, but it cannot be equal to $h_{ij}$, so no product of the entries below $h_{ij}$ can be equal to $h_{ij}$. Thus $g_i$ cannot be written as the product of any subset of the $g_j$ for $j > i$.

But this means that $g_i$ cannot be written as a product of any subset of the other $g_j$ for $j \neq i$, as we can prove by contradiction: suppose $g_i$ could be written as a product

$$g_i = \prod_{j \in \mathcal{J}} g_j, \tag{C8}$$

where $\mathcal{J}$ is some set of indices not including $i$. If $i < j$ for all $j \in \mathcal{J}$, then this violates the condition in the previous paragraph directly. Otherwise, choose the least $j \in \mathcal{J}$: call it $j'$. Then since the Pauli operators are self-inverse, and all of the $g_k$ commute, we can rearrange (C8) as

$$g_{j'} = g_i \prod_{j \in \mathcal{J} \setminus \{j'\}} g_j, \tag{C9}$$

which violates the condition in the previous paragraph since $j' < i, j$ for all $j \in \mathcal{J} \setminus \{j'\}$. Thus, all of the $g_i$ are independent.

We wish to be able to recover the original set of operators $G'$ from the new independent set $G$. To accomplish this, we simply need to keep track of the row multiplications carried out in each step above. Since the Pauli operators are self-inverse, we can reconstruct each row $i$ in $G'$ as the product of the corresponding row in $G$ and the rows it is multiplied by in the procedure (including in the multiplication the corresponding signs in $\vec{s}$). Thus we can write each operator in $G'$ as a product of operators in $G$.

## APPENDIX D: MORE GENERAL QUASIQUANTIZED MODELS

In Sec. II A we showed how to construct the ontic states (phase space points) for a noncontextual set of observables. The key step is obtaining the set $\mathcal{R} = \{C_{i1} \mid i = 1, 2, \ldots, N\} \cup G$, where $\{C_{i1} \mid i = 1, 2, \ldots, N\}$ is a completely anticommuting set, $G = \{G_i \mid i = 1, 2, \ldots, |G|\}$ is an independent commuting set whose elements also commute with each $C_{i1}$, and every operator in $\mathcal{S}$ may be written as a product of commuting operators in $\mathcal{R}$. Specifically, each universally commuting operator in $\mathcal{S}$ (these compose the set $\mathcal{Z}$) may be written as a product of operators in $G$, and every operator in one of the cliques $C_i \subset \mathcal{T}$ may be written as a product of operators in $G$ with the single operator $C_{i1}$.

The joint outcome assignments for $\mathcal{R}$ label the phase space points (each of which defines a joint outcome assignment for all of $\mathcal{S}$), so they also label the joint probabilities associated with each phase space point. As in Sec. II B, we write these joint probabilities as

$$P(c_1, \ldots, c_N, g_1, g_2, \ldots), \tag{D1}$$

where each $c_i, g_i$ is $\pm 1$ and denotes the outcome assigned to $C_{i1}$ or $G_i$, respectively.

In Sec. II B we presented a quasiquantized model that gives a set of joint probabilities $P$ (by way of a set of expectation values) that is sufficiently general to reproduce the expectation values associated with any eigenstate of the Hamiltonian. However, the set of ontic states in principle admits broader sets of probability distributions, which we discuss in this Appendix.

The probabilities for outcome $+1$ for each operator in $\mathcal{S}$ should be obtained as marginals of the joint probability distribution (D1). Consider an operator in $\mathcal{S}$ that is written as

$$B \text{ or } C_{i1}B, \tag{D2}$$

for $B$ a product of operators in $G$ (which, it is understood, may be the identity—the product of no operators in $G$). The corresponding marginal is the sum of the joint probabilities for outcomes such that the product outcome from (D2) is $+1$. In other words, if $p_O$ denotes the probability of obtaining outcome $+1$ upon measurement of the operator $O$, then

$$p_B = \sum_{\substack{c_i, g_j = \pm 1, \\ \text{s.t. } B \mapsto 1}} P(c_1, \ldots, c_N, g_1, g_2, \ldots), \tag{D3}$$

$$p_{C_{i1}B} = \sum_{\substack{c_i, g_j = \pm 1, \\ \text{s.t. } C_{i1}B \mapsto 1}} P(c_1, \ldots, c_N, g_1, g_2, \ldots). \tag{D4}$$

Exactly half of the joint probabilities will appear in each sum (D3) and (D4). This is apparent when we note that for any point $(c_1, \ldots, c_N, g_1, g_2, \ldots)$ satisfying, say, the condition $B = 1$ for $B$ a product of some subset of the $g_j$, flipping the sign of any of the coordinates $g_j$ appearing in the product will cause it to violate the condition, and vice versa.

Alternatively, we may visualize the sets of phase space points appearing in the sums (D3) and (D4) by first reassigning the outcome labels as

$$c_i \to c_i' = \tfrac{1}{2}(1 - c_i), \quad g_i \to g_i' = \tfrac{1}{2}(1 - g_i), \tag{D5}$$

i.e., the outcome assignments are mapped as

$$1 \to 0, \quad -1 \to 1, \tag{D6}$$

and products of outcomes become binary sums of outcomes. In other words, we have mapped our outcome space for each observable to $\mathbb{Z}_2$ (also denoted $\mathbb{F}_2$, the field of two elements), as is common practice in the literature (see [70], for example). The marginalizations (D3) and (D4) therefore become sums over all the joint probabilities for phase space points satisfyingconditions of the form

$$\vec{A} \cdot \vec{c}\,' + \vec{B} \cdot \vec{g}\,' = 1, \tag{D7}$$

where $\vec{c}\,' \equiv (c_1', \ldots, c_N')$ and $\vec{g}\,' \equiv (g_1', g_2', \ldots)$, and $\vec{A}, \vec{B}$ are vectors in $(\mathbb{Z}_2)^N$, $(\mathbb{Z}_2)^{|G|}$, respectively. We can now see that (D7) is the equation for a hyperplane in the phase space $(\mathbb{Z}_2)^{N+|G|}$, so we may think of the marginal probabilities for

TABLE V. $BeH_2$ Hamiltonian in Kandala *et al.*, 2017 [10]. Terms are displayed in the format {Pauli operator: coefficient,...}, i.e., as a Python dict mapping Pauli operators that appear in the Hamiltonian to their coefficients. Coefficients are given in Hartree.

**Diagonal terms:** {ZIIII: −0.143021, ZZIII: 0.104962, IZZII: 0.038195, IIZII: −0.325651, IIIZII: −0.143021, IIIZZI: 0.104962, IIIIZZ: 0.038195, IIIIIZ: −0.325651, IZIII: 0.172191, ZZZII: 0.174763, ZIZII: 0.136055, ZIIZII: 0.116134, ZIIZZI: 0.094064, ZIIIZZ: 0.099152, ZIIIIZ: 0.123367, ZZIZII: 0.094064, ZZIZZI: 0.098003, ZZIIZZ: 0.102525, ZZIIIZ: 0.097795, IZZZII: 0.099152, IZZZZI: 0.102525, IZZIZZ: 0.112045, IZZIIZ: 0.105708, IIZZII: 0.123367, IIZZZI: 0.097795, IIZIZZ: 0.105708, IIZIIZ: 0.133557, IIIIZI: 0.172191, IIIZZZ: 0.174763, IIIZIZ: 0.136055}

**Additional terms in noncontextual sub-Hamiltonian:** {IIZIZX: 0.010064, IIZIIX: −0.010064, ZIIIIX: −0.009922, ZIIIZX: 0.009922, IZZIIX: 0.007952, IZZIZX: −0.007952, ZZIIIX: 0.007016, ZZIIZX: −0.007016, IIIZZX: −0.002246, IIIZIX: 0.002246}

**Additional terms in full Hamiltonian:** {XZIIII: 0.059110, XIIIII: −0.059110, IZXIII: 0.161019, IIXIII: −0.161019, IIIXZI: 0.059110, IIIXII: −0.059110, IIIIZX: 0.161019, IIIIIX: −0.161019, XIXIII: −0.038098, XZXIII: −0.003300, XZIXZI: 0.013745, XZIXII: −0.013745, XIIXZI: −0.013745, XIIXII: 0.013745, XZIIZX: 0.011986, XZIIIX: −0.011986, XIIIZX: −0.011986, XIIIIX: 0.011986, IZXXZI: 0.011986, IZXXII: −0.011986, IIXXZI: −0.011986, IIXXII: 0.011986, IZXIZX: 0.013836, IZXIIX: −0.013836, IIXIZX: −0.013836, IIXIIX: 0.013836, IIIXIX: −0.038098, IIIXZX: −0.003300, ZZXIII: −0.002246, ZIXIII: 0.002246, ZIIXZI: 0.014815, ZIIXII: −0.014815, ZZIXZI: −0.002038, ZZIXII: 0.002038, XIZIII: −0.006154, XZZIII: 0.006154, XZIZII: 0.014815, XIIZII: −0.014815, XZIZZI: −0.002038, XIIZZI: 0.002038, XZIIZZ: 0.001124, XIIIZZ: −0.001124, XZIIIZ: 0.017678, XIIIIZ: −0.017678, YIYIII: −0.041398, YYIXXZ: 0.011583, YYIIXI: −0.011094, IYYXXZ: 0.010336, IYYIXI: −0.005725, IIIXIZ: −0.006154, XXZXXZ: 0.011583, XXZIXI: −0.011094, IXIXXZ: −0.011094, IXIIXI: 0.026631, IIZXII: −0.017678, XXZYYI: 0.011583, XXZIYY: 0.010336, IXIYYI: −0.011094, IXIIYY: −0.005725, IIIYIY: −0.041398, YYIYYI: 0.011583, YYIIYY: 0.010336, IYYYYI: 0.010336, IYYIYY: 0.010600, XXZXXX: 0.024909, IXIXXX: −0.031035, XXZYXY: 0.024909, IXIYXY: −0.031035, YYIXXX: 0.024909, IYYXXX: 0.021494, YYIYXY: 0.024909, IYYYXY: 0.021494, XXZZXZ: 0.011094, IXIZXZ: −0.026631, YYIZXZ: 0.011094, IYYZXZ: 0.005725, XXZZXX: 0.010336, IXIZXX: −0.005725, YYIZXX: 0.010336, IYYZXX: 0.010600, XXXXXZ: 0.024909, XXXIXI: −0.031035, IIXIIZ: −0.010064, XXXYYI: 0.024909, XXXIYY: 0.021494, YXYXXZ: 0.024909, YXYIXI: −0.031035, YXYYYI: 0.024909, YXYIYY: 0.021494, XXXXXX: 0.063207, XXXYXY: 0.063207, YXYXXX: 0.063207, YXYYXY: 0.063207, XXXZXZ: 0.031035, IIXZII: −0.009922, YXYZXZ: 0.031035, XXXZXX: 0.021494, YXYZXX: 0.021494, ZXZXXZ: 0.011094, ZXZIXI: −0.026631, ZXZYYI: 0.011094, ZXZIYY: 0.005725, ZXZXXX: 0.031035, ZXZYXY: 0.031035, ZXZZXZ: 0.026631, ZXZZXX: 0.005725, ZXXXXZ: 0.010336, ZXXIXI: −0.005725, ZXXYYI: 0.010336, ZXXIYY: 0.010600, ZXXXXX: 0.021494, ZXXYXY: 0.021494, ZXXZXZ: 0.005725, ZXXZXX: 0.010600, IZZXZI: 0.001124, IZZXII: −0.001124, IIZXZI: 0.017678, IZXZII: 0.009922, IZXZZI: −0.007016, IIXZZI: 0.007016, IZXIZZ: −0.007952, IIXIZZ: 0.007952, IZXIIZ: 0.010064, IIIXZZ: 0.006154}

**Minimal energy parameter setting for $\mathcal{R}$, in noncontextual sub-Hamiltonian:**

$q_1 = \langle ZZZIII\rangle \mapsto -1, \quad q_2 = \langle IIIZII\rangle \mapsto +1, \quad q_3 = \langle IZIIII\rangle \mapsto -1,$

$q_4 = \langle IIIIZI\rangle \mapsto -1, \quad q_5 = \langle ZIIIII\rangle \mapsto +1,$

$r_1 = \langle IIIIZX\rangle \mapsto -0.7522001251805058, \quad r_2 = \langle IIIIIZ\rangle \mapsto 0.6589347248995392$

outcomes of individual measurements as sums of the joint probabilities over such hyperplanes [70].

We wish to demonstrate that for any state there exists a joint probability distribution $P$ that reproduces as marginals the correct probabilities for the outcomes of any observable in $\mathcal{S}$. Since each ontic state carries with it an outcome assignment to each observable in $\mathcal{S}$, it is enough to show that for any state, $P$ can reproduce as marginals the correct probabilities for the outcomes of any observable in $\mathcal{R}$. The largest subsets of $\mathcal{R}$ that may be measured simultaneously are $\{C_{i1}\} \cup G$ for any $i = 1, 2, \ldots, N$: let us refer to the probabilities for the joint outcomes for these as

$$P^{(\{C_{i1}\}\cup G)}(c_i, g_1, g_2, \ldots), \tag{D8}$$

where the $c_i, g_1, g_2, \cdots = \pm 1$ label the joint outcomes. Each $C_{i1}$ commutes with all operators in $G$, so for a given state we can directly determine (via Born's rule, using the appropriate projectors, or by actual measurements if the state is physical) the probabilities (D8) (separately for each $i$). $P$ will thus correctly reproduce as marginals the probabilities for the observables in $\mathcal{R}$ if and only if it correctly reproduces as marginals the joint outcome probabilities (D8), since these correspond to the largest simultaneously measurable subsets of $\mathcal{R}$.

Thus, we wish $P$ to satisfy

$$P^{(\{C_{i1}\}\cup G)}(c_i, g_1, g_2, \ldots) = \sum_{\substack{c_j = \pm 1, \\ \forall j \neq i}} P(c_1, \ldots, c_i, \ldots, c_N, g_1, g_2, \ldots) \tag{D9}$$

for each $i = 1, 2, \ldots, N$. Since the $C_{i1}$ pairwise anticommute, the probabilities $P^{(\{C_{i1}\}\cup G)}$ and $P^{(\{C_{j1}\}\cup G)}$ for any $i \neq j$ cannot be correlated. More specifically, the product $C_{i1}B_iC_{j1}B_j$ of the operators $C_{i1}B_i$ and $C_{j1}B_j$ (for $i \neq j$ and $B_i$, $B_j$ any products of operators in $G$) is not jointly measurable with $C_{i1}B_i$ and $C_{j1}B_j$, since none of the three commute with each other, and in general the product need not even be in $\overline{\mathcal{S}}$. In any case, it has no bearing on the joint probabilities for $C_{i1}B_i$ and $C_{j1}B_j$; all that is required of the joint probability distribution $P$ is that it correctly reproduce the expressions (D9), which the assignment (D10) does. Therefore, we may take the probability for the

joint outcome $P(c_1, \ldots, c_N, g_1, g_2, \ldots)$ to be proportional to the product of the probabilities $P^{(\{C_{i1}\}\cup G)}(c_i, g_1, g_2, \ldots)$ over all $i = 1, 2, \ldots, N$. Including the appropriate normalization gives an expression for the joint probabilities:

$$\frac{P(c_1, \ldots, c_N, g_1, g_2, \ldots)}{P^{(G)}(g_1, g_2, \ldots)} = \prod_{i=1}^{N} \frac{P^{(\{C_{i1}\}\cup G)}(c_i, g_1, g_2, \ldots)}{P^{(G)}(g_1, g_2, \ldots)}, \tag{D10}$$

where $P^{(G)}(g_1, g_2, \ldots)$ denotes the probability of obtaining the joint outcome $(g_1, g_2, \ldots)$ for the operators $G$ (which we can also obtain via Born's rule).

As we noted above, given any quantum state we can directly evaluate each of the probabilities on the right-hand side, since each is the probability for a joint outcome of one of the commuting sets of observables $\{C_{i1}\} \cup G$ (or just $G$). Therefore, for any quantum state, (D10) gives a joint probability distribution $P(c_1, \ldots, c_N, g_1, g_2, \ldots)$ for the phase space points $(c_1, \ldots, c_N, g_1, g_2, \ldots)$, which reproduces as marginals all probabilities associated with the state, for observables in $\mathcal{S}$.

This discussion does not provide specific methods for constructing quasiquantized models alternative to that presented in the main text. What we have shown, however, is that epistemic states over the ontic states described in Sec. II A may in principle describe the measurable properties of $\mathcal{S}$ for any state. Thus, the door is open for broader sets of epistemic states than those allowed in our model as described in Sec. II B. For our purposes, however, the set of epistemic states allowed in our model is sufficient.

## APPENDIX E: NONCONTEXTUAL SUB-HAMILTONIANS

For the $HeH^+$ Hamiltonian in [2], we performed a brute-force search over all noncontextual sub-Hamiltonians to find the one that gave the best approximation to the full Hamiltonian's ground state energy.

For the other three Hamiltonians in Table I, as described in Sec. IV we used a greedy heuristic to find large noncontextual sub-Hamiltonians. This "one-by-one" heuristic greedily selected terms from the full Hamiltonian, in decreasing order of weight, while the set remained noncontextual. In Tables II–V we list the full Hamiltonians together with the best noncontextual sub-Hamiltonians we found, as well as the diagonal terms only. We also tried greedily adding larger subsets up to subsets of size six (four, in the $BeH_2$ Hamiltonian of [10]), as well as checking for noncontextual subsets with certain fixed structures (by brute-force search over generating sets $\mathcal{R}$ where the anticommuting generators $C_{i1}$ act on fixed subsets of the qubits). We found that none of the results outperformed the one-by-one greedy approach, and those that performed best simply reproduced the same noncontextual sub-Hamiltonian as the one-by-one greedy approach.

We obtained the ground state energies for the full Hamiltonians and diagonal sub-Hamiltonians by evaluating them directly using the OpenFermion software package [71]. For the noncontextual sub-Hamiltonians, we evaluated the ground state energies using our quasiquantized model as described in Sec. III A, and optimizing by a brute-force search over the parameter space $E$. We then checked the resulting ground state energies against those computed by OpenFermion, and found agreement to machine precision in all cases.


[1] J. Preskill, Quantum **2**, 79 (2018).

[2] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'Brien, Nat. Commun. **5**, 4213 (2014).

[3] J. Du, N. Xu, X. Peng, P. Wang, S. Wu, and D. Lu, Phys. Rev. Lett. **104**, 030502 (2010).

[4] B. P. Lanyon, J. D. Whitfield, G. G. Gillett, M. E. Goggin, M. P. Almeida, I. Kassal, J. D. Biamonte, M. Mohseni, B. J. Powell, M. Barbieri, A. Aspuru-Guzik, and A. G. White, Nat. Chem. **2**, 106 (2010).

[5] Y. Wang, F. Dolde, J. Biamonte, R. Babbush, V. Bergholm, S. Yang, I. Jakobi, P. Neumann, A. Aspuru-Guzik, J. D. Whitfield, and J. Wrachtrup, ACS Nano **9**, 7769 (2015).

[6] P. J. J. O'Malley, R. Babbush, I. D. Kivlichan, J. Romero, J. R. McClean, R. Barends, J. Kelly, P. Roushan, A. Tranter, N. Ding, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, A. G. Fowler, E. Jeffrey, E. Lucero, A. Megrant, J. Y. Mutus, M. Neeley, C. Neill, C. Quintana, D. Sank, A. Vainsencher, J. Wenner, T. C. White, P. V. Coveney, P. J. Love, H. Neven, A. Aspuru-Guzik, and J. M. Martinis, Phys. Rev. X **6**, 031007 (2016).

[7] R. Santagati, J. Wang, A. A. Gentile, S. Paesani, N. Wiebe, J. R. McClean, S. Morley-Short, P. J. Shadbolt, D. Bonneau, J. W. Silverstone, D. P. Tew, X. Zhou, J. L. O'Brien, and M. G. Thompson, Sci. Adv. **4**, eaap9646 (2018).

[8] Y. Shen, X. Zhang, S. Zhang, J.-N. Zhang, M.-H. Yung, and K. Kim, Phys. Rev. A **95**, 020501(R) (2017).

[9] S. Paesani, A. A. Gentile, R. Santagati, J. Wang, N. Wiebe, D. P. Tew, J. L. O'Brien, and M. G. Thompson, Phys. Rev. Lett. **118**, 100503 (2017).

[10] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, Nature (London) **549**, 242 (2017).

[11] C. Hempel, C. Maier, J. Romero, J. McClean, T. Monz, H. Shen, P. Jurcevic, B. P. Lanyon, P. Love, R. Babbush, A. Aspuru-Guzik, R. Blatt, and C. F. Roos, Phys. Rev. X **8**, 031022 (2018).

[12] E. F. Dumitrescu, A. J. McCaskey, G. Hagen, G. R. Jansen, T. D. Morris, T. Papenbrock, R. C. Pooser, D. J. Dean, and P. Lougovski, Phys. Rev. Lett. **120**, 210501 (2018).

[13] J. I. Colless, V. V. Ramasesh, D. Dahlen, M. S. Blok, M. E. Kimchi-Schwartz, J. R. McClean, J. Carter, W. A. de Jong, and I. Siddiqi, Phys. Rev. X **8**, 011021 (2018).

[14] Y. Nam, J.-S. Chen, N. C. Pisenti, K. Wright, C. Delaney, D. Maslov, K. R. Brown, S. Allen, J. M. Amini, J. Apisdorf, K. M. Beck, A. Blinov, V. Chaplin, M. Chmielewski, C. Collins, S. Debnath, A. M. Ducore, K. M. Hudek, M. Keesan, S. M. Kreikemeier, J. Mizrahi, P. Solomon, M. Williams, J. D. Wong-Campos, C. Monroe, and J. Kim, arXiv:1902.10171.

[15] C. Kokail, C. Maier, R. van Bijnen, T. Brydges, M. K. Joshi, P. Jurcevic, C. A. Muschik, P. Silvi, R. Blatt, C. F. Roos, and P. Zoller, Nature (London) **569**, 355 (2019).

[16] A. Kandala, K. Temme, A. D. Córcoles, A. Mezzacapo, J. M. Chow, and J. M. Gambetta, Nature (London) **567**, 491 (2019).
[17] W. M. Kirby and P. J. Love, Phys. Rev. Lett. **123**, 200501 (2019).
[18] J. S. Bell, Physics **1**, 195 (1964).
[19] J. S. Bell, Rev. Mod. Phys. **38**, 447 (1966).
[20] S. Kochen and E. Specker, J. Math. Mech. **17**, 59 (1967).
[21] R. W. Spekkens, Phys. Rev. Lett. **101**, 020401 (2008).
[22] C. Ferrie and J. Emerson, J. Phys. A: Math. Theor. **41**, 352001 (2008).
[23] C. Ferrie and J. Emerson, New J. Phys. **11**, 063040 (2009).
[24] C. Ferrie, Rep. Prog. Phys. **74**, 116001 (2011).
[25] V. Veitch, C. Ferrie, D. Gross, and J. Emerson, New J. Phys. **14**, 113011 (2012).
[26] R. Raussendorf, D. E. Browne, N. Delfosse, C. Okay, and J. Bermejo-Vega, Phys. Rev. A **95**, 052334 (2017).
[27] R. Raussendorf, Phys. Rev. A **88**, 022322 (2013).
[28] M. Howard, J. Wallman, V. Veitch, and J. Emerson, Nature (London) **510**, 351 (2014).
[29] A. Karanjai, J. J. Wallman, and S. D. Bartlett, arXiv:1802.07744.
[30] R. Raussendorf, Quantum Inf. Comput. **19**, 1141 (2019).
[31] D. Schmid, R. W. Spekkens, and E. Wolfe, Phys. Rev. A **97**, 062103 (2018).
[32] C. Duarte and B. Amaral, J. Math. Phys. **59**, 062202 (2018).
[33] C. Okay, E. Tyhurst, and R. Raussendorf, Quantum Inf. Comput. **18**, 1272 (2018).
[34] M. Frembs, S. Roberts, and S. D. Bartlett, New J. Phys. **20**, 103011 (2018).
[35] D. R. M. Arvidsson-Shukur, N. Y. Halpern, H. V. Lepage, A. A. Lasek, C. H. W. Barnes, and S. Lloyd, Nat Commun. **11**, 3775 (2020).
[36] S. Abramsky and A. Brandenburger, New J. Phys. **13**, 113036 (2011).
[37] R. Ramanathan, A. Soeda, P. Kurzyński, and D. Kaszlikowski, Phys. Rev. Lett. **109**, 050404 (2012).
[38] A. Cabello, S. Severini, and A. Winter, Phys. Rev. Lett. **112**, 040401 (2014).
[39] R. Ramanathan and P. Horodecki, Phys. Rev. Lett. **112**, 040404 (2014).
[40] A. Grudka, K. Horodecki, M. Horodecki, P. Horodecki, R. Horodecki, P. Joshi, W. Kłobus, and A. Wójcik, Phys. Rev. Lett. **112**, 120401 (2014).
[41] A. Cabello, M. Kleinmann, and C. Budroni, Phys. Rev. Lett. **114**, 250402 (2015).
[42] S. Abramsky, R. S. Barbosa, and S. Mansfield, Phys. Rev. Lett. **119**, 050504 (2017).
[43] N. de Silva, Phys. Rev. A **95**, 032108 (2017).
[44] B. Amaral and M. T. Cunha, arXiv:1709.04812.
[45] Z.-P. Xu and A. Cabello, Phys. Rev. A **99**, 020103(R) (2019).
[46] S. Mansfield and E. Kashefi, Phys. Rev. Lett. **121**, 230401 (2018).
[47] R. Raussendorf, J. Bermejo-Vega, E. Tyhurst, C. Okay, and M. Zurel, Phys. Rev. A **101**, 012350 (2020).
[48] D. Schmid, J. Selby, E. Wolfe, R. Kunjwal, and R. W. Spekkens, arXiv:1911.10386.
[49] R. W. Spekkens, Quasi-quantization: Classical statistical theories with an epistemic restriction, in *Quantum Theory: Informational Foundations and Foils*, edited by G. Chiribella and R. W. Spekkens (Springer Netherlands, Dordrecht, 2016), pp. 83–135.
[50] R. W. Spekkens, Phys. Rev. A **75**, 032110 (2007).
[51] S. J. van Enk, Found. Phys. **37**, 1447 (2007).
[52] S. D. Bartlett, T. Rudolph, and R. W. Spekkens, Phys. Rev. A **86**, 012103 (2012).
[53] F. Barahona, J. Phys. A: Math. Gen. **15**, 3241 (1982).
[54] A. Y. Kitaev, A. H. Shen, and M. N. Vyalyi, *Classical and Quantum Computation, Volume 47 of Graduate Studies in Mathematics* (American Mathematical Society, Providence, RI, 2002).
[55] P. Wocjan and T. Beth, Int. J. Quantum Inf. **01**, 349 (2003).
[56] J. Kempe and O. Regev, Quantum Inf. Comput. **3**, 258 (2003).
[57] S. Bravyi and M. Vyalyi, Quantum Inf. Comput. **5**, 187 (2005).
[58] J. Kempe, A. Kitaev, and O. Regev, SIAM J. Comput. **35**, 1070 (2006).
[59] R. Oliveira and B. M. Terhal, Quantum Inf. Comput. **8**, 0900 (2008).
[60] J. D. Biamonte and P. J. Love, Phys. Rev. A **78**, 012352 (2008).
[61] D. Aharonov and L. Eldar, in *2011 IEEE 52nd Annual Symposium on Foundations of Computer Science* (IEEE, Piscataway, NJ, 2011), pp. 334–343.
[62] N. Schuch, Quantum Inf. Comput. **11**, 901 (2011).
[63] J. Yan and D. Bacon, arXiv:1203.3906.
[64] D. Aharonov and L. Eldar, Quant. Info. Proc. **14**, 83 (2015).
[65] S. McArdle, S. Endo, A. Aspuru-Guzik, S. Benjamin, and X. Yuan, Rev. Mod. Phys. **92**, 015003 (2020).
[66] Note that this problem is distinct from that shown to be in P in [63], in which the Hamiltonian is a linear combination of commuting Pauli operators *with coefficients* $\pm 1$.
[67] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, UK, 2001).
[68] K. Jansen, P. Scheffler, and G. Woeginger, in *Proceedings of the Fundamentals of Computation Theory*, Lecture Notes in Computer Science (Springer, 1993), Vol. 710, pp. 319–328.
[69] E. Farhi, J. Goldstone, and S. Gutmann, arXiv:1411.4028.
[70] K. S. Gibbons, M. J. Hoffman, and W. K. Wootters, Phys. Rev. A **70**, 062101 (2004).
[71] J. R. McClean, N. C. Rubin, K. J. Sung, I. D. Kivlichan, X. Bonet-Monroig, Y. Cao, C. Dai, E. S. Fried, C. Gidney, B. Gimby, P. Gokhale, T. Häner, T. Hardikar, V. Havlíček, O. Higgott, C. Huang, J. Izaac, Z. Jiang, X. Liu, S. McArdle, M. Neeley, T. O'Brien, B. O'Gorman, I. Ozfidan, M. D. Radin, J. Romero, N. P. D. Sawaya, B. Senjean, K. Setia, S. Sim, D. S. Steiger, M. Steudtner, Q. Sun, W. Sun, D. Wang, F. Zhang, and R. Babbush, Quantum Sci. Tech. **5**, 034014 (2020).