Shot-Noise Reduction for Lattice Hamiltonians

Timo Eckstein,[1,2,*] Refik Mansuroglu,[1,3] Stefan Wolf,[1] Ludwig Nützel,[1] Stephan Tasler,[1] Martin Kliesch,[4] and Michael J. Hartmann[1,2]

[1] *Department of Physics, Friedrich-Alexander Universität Erlangen-Nürnberg, Erlangen, Germany*
[2] *Max-Planck Institute for the Science of Light, Erlangen, Germany*
[3] *Faculty of Physics, University of Vienna, Vienna, Austria*
[4] *Hamburg University of Technology, Institute for Quantum Inspired and Quantum Optimization, Hamburg, Germany*

(Received 14 November 2024; revised 15 July 2025; accepted 14 January 2026; published 3 April 2026)

Efficiently estimating energy expectation values of quantum lattice systems on quantum computers is a crucial subroutine for various quantum algorithms, which can lead to significant overhead due to the high measurement shot numbers required. We introduce a measurement strategy tailored to quantum lattice systems and (noisy) energy eigenstates. It is based on a geometric partitioning of the Hamiltonian into local patches and performing the measurements in the eigenbases of those patches. The resulting energy estimator has a smaller variance than the ones of Pauli grouping schemes, which leads to a reduction of the total number of shots. We provide rigorous guarantees for this variance improvement for energy eigenstates, also in the presence of depolarizing noise. As one can choose the subsystem size, one can ensure that measurement circuits remain within implementable depths. In numerical experiments, we demonstrate the shot count reduction for various 2D lattice models, including the transverse field XY and Ising models, as well as the Fermi-Hubbard model. We find sampling improvements of several orders of magnitude already for plaquettes of two by two qubits, where the required readout circuits remain very moderate in depth.

DOI: 10.1103/xy36-drb3

I. INTRODUCTION

Extracting information from a quantum computer means sampling from a probability distribution defined by the prepared quantum state and the observable to be measured. Since the speed of individual gate and readout operations for quantum computers is much slower than for their classical counterparts, the efficiency of a quantum algorithm or quantum simulation is often limited by the number of measurements and hence executions that need to be made to obtain sufficient samples. Therefore, reducing the sampling complexity in obtaining the output of a quantum computer is a goal of utmost importance.

Since quantum computers typically can only perform measurements in the computational basis, the observable of interest needs to be split into parts that can be measured individually after rotating their eigenbases into the computational basis. Thereby, the individual parts typically only comprise single- or two-qubit strings of Pauli operators, where the required basis rotations can be done via single-qubit gates.

The efficiency of measurement strategies in quantum computation is a very active field of research [1–3], where the challenge of reducing shot counts has received significant attention, for example, in the context of applications in quantum chemistry. For corresponding measurement strategies, Pauli grouping methods [4–9], on the one hand, aim to reduce the number of parts that the observable is split into, which corresponds to minimizing the state-independent worst-case measurement error. Classical shadows [10–14], on the other hand, aim at estimating expectation values in an observable-independent fashion, meaning one is not restricted to knowing *a priori* a specific observable of interest, but also does not adapt the sampling strategy to it. Instead, one aims to use the locality of the to-be-estimated observables so that local random measurements suffice to reduce the measurement effort [8,15]. Here, in contrast, we use the structure of both the quantum state and the observable to reduce measurement uncertainty.

Besides quantum chemistry applications, the investigation of condensed matter quantum many-body systems is

*Contact author: Timo.Eckstein@fau.de

also considered a very important and promising application for quantum computers. Moreover, the Hamiltonians of such systems can often be natively explored on solid-state quantum computing platforms [16–19] with a fixed connectivity graph. While the Hamiltonians from quantum chemistry applications have a complex nonlocal interaction structure, condensed matter systems often have the form of quantum lattice models, where grouping methods only have little effect on the measurement count. Hence, different strategies are required for obtaining significant improvements in the required number of measurements for lattice systems, a challenge that has received relatively little attention so far.

In this work, we tailor measurement schemes for this broad class of quantum lattice models by splitting the observable of interest, for example, a Hamiltonian, into parts that have support on multiple adjacent lattice sites and measure these parts in their eigenbasis. We call this *geometric partitioning*. When applying it to a system that is prepared in an eigenstate of the total Hamiltonian one may expect that it leads to reduced sampling requirements.

The intuition behind this expectation is the following. For a Hamiltonian with short-range interactions, the expansion of the global energy eigenstate in a basis formed by products of eigenstates of contiguous multi-qubit parts of the same Hamiltonian (geometric partitioning) should contain fewer terms than an expansion in terms of products of eigenstates of individual qubits, which one may call Pauli partitioning. The reason is that in geometric partitioning, all the Hamiltonian terms except for the interactions between patches are diagonal, whereas all qubit-qubit interactions (which are many more) are not diagonalized in Pauli partitioning. In this work, we quantify this sampling advantage in terms of the variances of the employed estimators, which allows us to derive lower bounds on the sampling improvement. We also show in numerical simulations that the improvement can often be several orders of magnitude already for measurements in the eigenbases of plaquettes of two by two qubits, where the required readout circuits remain very moderate in depth.

Our work has the following main results. In Theorem 1, we show that, for systems in an energy eigenstate, geometric partitions always outperform Pauli partitions in terms of the required sample number for eigenenergy estimation and obtain lower bounds for the improvement factors. We then numerically show reductions of sampling complexity of several orders of magnitude for a number of quantum lattice models as examples. Importantly, these reductions are already seen for partitionings into very moderate-sized parts, where the measurement circuits can feasibly be implemented.

In Theorem 2, we then show that sampling improvements carry over to imperfect eigenstates, which one expects from preparation circuits on real devices. As corollaries, we also obtain upper bounds on the extra noise that the measurement circuits are allowed to generate, see Corollary 2, and the number of gates they may be composed of, see Corollary 3, while still leading to a sampling improvement. These findings are again corroborated by numerical studies for several models.

The remainder of the paper is organized as follows. In Sec. II we explain the general concept of our approach, before deriving analytical bounds on the sampling improvement for exact eigenstates and comparing these to numerical tests in Sec. III. Section IV generalizes the analytical results to eigenstates perturbed by global depolarizing noise, including numerical examples. Finally, Sec. V provides a summary and conclusions.

II. CONCEPT

We first present the concept behind our approach. In our discussion, we consider the energy of a Hamiltonian H and its exact or approximate eigenstates as the objects of interest. Our results, however, equally apply to other observables with a lattice structure and their exact or approximate eigenstates.

Measuring the energy H for a quantum state $|\psi\rangle$ means randomly drawing the eigenvalues E_m of H with probabilities defined by the amplitudes' squares $|c_i|^2$ of the expansion $|\psi\rangle = \sum_i c_i |E_i\rangle$ in the eigenbasis $\{|E_i\rangle\}$ of H. The empirical mean estimator $\overline{H}_M := M^{-1} \sum_{m=1}^{M} E_m$ from M i.i.d. repetitions (or samples) then yields an approximation to the expectation value $\langle H\rangle = \langle\psi|H|\psi\rangle$ with a standard error σ_H given by

$$\sigma_H^2 = \mathrm{Var}\left(\overline{H}_M\right) = \frac{1}{M}\,\mathrm{Var}_{|\psi\rangle}(H). \tag{1}$$

This variance allows one to quantify finite sampling noise since Chebyshev's inequality upper bounds the probability to be more than a threshold ϵ away from the true expectation value,

$$P\left(\left|\overline{H}_M - \langle H\rangle\right| \geq \epsilon\right) \leq \mathrm{Var}\left(\overline{H}_M\right)/\epsilon^2. \tag{2}$$

Since measurements in quantum computation can typically only be done in the Pauli Z (or computational) basis, measuring an observable that is not diagonal in this basis would generally require a diagonalization $U_H H U_H^\dagger$, which is equivalent to implementing a rotation U_H that transforms the eigenbasis of H into the computational basis at the end of the quantum circuit. In general, however, this is a nontrivial task, as finding and implementing the basis transformation U_H is typically harder than solving the quantum computational problem in the first place.

If the eigenbasis of H cannot be obtained efficiently, it is possible to measure parts H_b, whose eigenbases are known, and that constitute the observable $H = \sum_b H_b$. The linearity of the expectation value then allows one to estimate

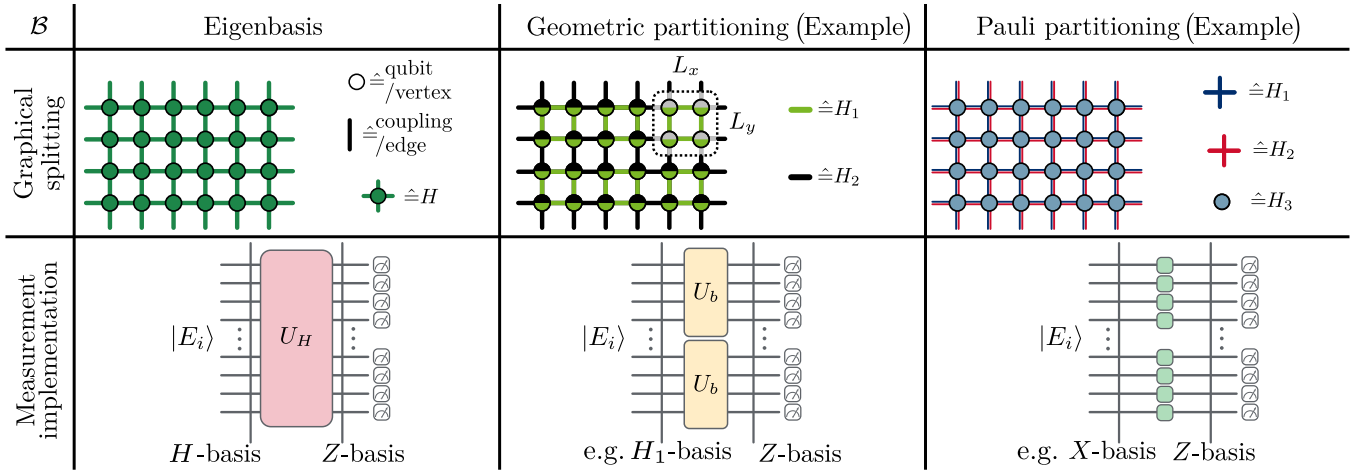

FIG. 1. Concept of geometric partitioning. For lattice Hamiltonians as in Eq. (5), one can measure the energy in the eigenbasis (first column), which requires implementing a measurement unitary corresponding to a full diagonalization of the Hamiltonian, or in Pauli bases (third column), which only requires single-qubit unitaries, as extreme cases. Here, we explore the regime between these extremes, namely measurement in the eigenbases of disjoint subsystems of H, which each comprise multiple lattice sites (second column). We call this a geometric partitioning. In the lattice sketch of the second column, this strategy is indicated by coloring complementary parts of the lattice in different colors (green and black). The idea is to measure all green parts in one shot and all black parts in another shot. In doing so, the single-site terms of the Hamiltonian are split into two identical parts to make the splitting symmetric, which is indicated by coloring the corresponding dots half in one and half in the other color. Geometric partitionings naturally take the local structure of the observable H into account while keeping the basis transformations $U_{H_b} = U_b \otimes \cdots \otimes U_b$ scalably implementable due to their product structure. As our main analytical results, we show in Theorem 1 that these geometric partitionings of H always lead to fewer required measurements for estimating E_i from $|E_i\rangle$ compared to the Pauli partitioning baseline for the same measurement uncertainty. Furthermore, these sampling improvements also apply to imperfect eigenstates.

$\langle H\rangle = \sum_b \langle H_b\rangle$. The empirical mean estimator of such a partition $\mathcal{B} = \{H_1, H_2, \ldots\}$ of H is given by

$$\overline{H}_M^{\mathcal{B}} := \sum_b \left(\frac{1}{M_b} \sum_{m=1}^{M_b} E_{b,m} \right), \tag{3}$$

where $E_{b,m}$ labels the sampled eigenvalues of the observable H_b, and M_b is the number of measurements for each H_b so that the overall number of measurements is $M = \sum_b M_b$. Since the individual samples $E_{b,m}$ are obtained in mutually independent repetitions of the experiment, the variance of the estimate $\overline{H}_M^{\mathcal{B}}$, is the sum of the variances of the empirical mean estimators for the individual H_b,

$$\tilde{\sigma}_{\mathcal{B}}^2 := \mathrm{Var}\left(\overline{H}_M^{\mathcal{B}}\right) = \sum_b \frac{\mathrm{Var}_{|\psi\rangle}(H_b)}{M_b}. \tag{4}$$

While the standard way of decomposing H into measurable parts H_b is to partition the Pauli strings in H into mutually commuting groups of Pauli sums, we show here that much more sampling-efficient measurement strategies can be obtained by making use of the structure of lattice Hamiltonians and their eigenstates.

While keeping much of our analysis general, we focus on translation-invariant Hamiltonians H with nearest-neighbor interactions on a rectangular lattice, as can be found in numerous quantum many-body systems, with the Ising or the Hubbard models as special cases,

$$H = \sum_i \mathcal{H}_i + \sum_{\langle i,j\rangle} \mathcal{H}_{i,j}, \tag{5}$$

where $\mathcal{H}_i = \sum_\alpha h_i^{(\alpha)} o_i^{(\alpha)}$ and $\mathcal{H}_{i,j} = \sum_{\alpha,\beta} J_{i,j}^{(\alpha,\beta)} o_i^{(\alpha)} o_j^{(\beta)}$ for the one-local and two-local terms. The Latin site indices i and j have two components for two-dimensional lattices, $i = (i_x, i_y)$ and $j = (j_x, j_y)$, whereas Greek indices label single operators $o_i^{(\alpha)}$ on individual lattice sites. These can be Pauli operators $o_i^{(\alpha)} \in \{\mathbb{1}, X, Y, Z\}$ for spin models, or creation, annihilation and number operators in Fock space representation $o_i^{(\alpha)} \in \{c_{i,\alpha}^\dagger, c_{i,\alpha}, n_{i,\alpha}\}$ for fermionic lattice models. The coefficients, $h_i^{(\alpha)}$ and $J_{i,j}^{(\alpha,\beta)}$ are real-valued coupling constants.

The main idea we pursue for realizing sampling improvement is to partition the geometrically local Hamiltonian H into terms H_b, which further decompose into subsystems $H_{b,k}$ of disjoint support,

$$H = \sum_b H_b \quad \text{with} \quad H_b = \sum_{k=1}^{K} H_{b,k}, \tag{6}$$

where the $H_{b,k}$ act on multiple adjacent lattice sites. The transformations of the measurement bases, from the eigenbases of the H_b into the computational basis, thus become direct products

$$U_{H_b} = U_{b,1} \otimes \cdots \otimes U_{b,K}; \tag{7}$$

see also Fig. 1. We call this a geometric partitioning and compare sample numbers M for different geometric partitions with the standard partition into fewest groups of mutually commuting local operators, $o_i^{(\alpha)}$ and $o_i^{(\alpha)} o_j^{(\beta)}$ in Eq. (5), under the condition of admitting the same standard error according to Eq. (4).

Geometric partitioning becomes particularly useful if the support of the individual $H_{b,k}$ grows beyond the interaction range, such that correlations of longer and longer range can be captured well. Nonetheless, the implementation of the transformations $U_{b,k}$, which is required in this strategy, can be done efficiently in many cases for the following reasons.

As one can choose the patch size for the $H_{b,k}$, it can be chosen constant in terms of system size n. For this choice, implementing U_{H_b} uses only $\mathcal{O}(n)$ gates, as implementing $U_{b,k}$ has constant gate complexity, and the number of $U_{b,k}$s to implement U_{H_b} scales only linearly in system size n. Moreover, the prepared state $|\psi\rangle = |E_i\rangle$ and local unitary U_b correspond to the same lattice Hamiltonian, but only differ in size. Thus, if there is an algorithm simplifying the state preparation, as, e.g., in Ref. [20], it could be used on a smaller scale, entry by entry, to simplify the implementation of U_b as well. In addition, for translationally invariant systems, all $U_{b,k} = U_b$ will be the same so that one needs to find only one such subsystem diagonalization circuit.

Furthermore, for specific diagonalizable 1D lattice models, like the transverse field Ising model, exact linear depth circuits to implement U_b are known [20] and can thus be used for partitionings that consider patch sizes $1 \times \sqrt{n}$. In addition, since the state $|\psi\rangle$ will typically not be prepared exactly, an approximation of U_b with a similar precision will be sufficient for our aims. Allowing an error ϵ in the implementation of U_b, one may thus make use of the Solovay-Kitaev algorithm [21,22], which has complexity $\mathcal{O}((\log(1/\epsilon))^c)$, $c \approx 1.44$ [23], and approximate it with a sequence of easier-to-implement unitaries.

To show the advantages of geometric partitioning, we first seek a quantitative measure to compare a measurement strategy $\mathcal{B}_1$ that splits $H = \sum_{H_b \in \mathcal{B}_1} H_b$ with another one $\mathcal{B}_2$ that splits $H = \sum_{H_b \in \mathcal{B}_2} H_b$ by determining which one requires less measurements for achieving the same standard error according to Eq. (4). To achieve this, we first need to resolve the issue that the standard error of a partitioned observable in Eq. (4) not only depends on the different partitionings but also on the specific measurement budget allocation M_b. It turns out that comparing partitioning strategies for the optimal measurement budget allocation, which minimizes the sampling error, makes the comparison independent of the distribution of the M_b [15],

$$\sigma_{\mathcal{B}}^2 = \min_{M_{\mathcal{B}} = \sum M_b} \tilde{\sigma}_{\mathcal{B}}^2 = \frac{1}{M_{\mathcal{B}}} \left[\sum_{H_b \in \mathcal{B}} \sqrt{\mathrm{Var}_{|\psi\rangle}(H_b)} \right]^2. \tag{8}$$

One can prove Eq. (8) using the positivity of the variance and minimizing $\sigma_{\mathcal{B}}^2$ using Lagrange multipliers to find that all $\sqrt{\mathrm{Var}_{|\psi\rangle}(H_b)}/M_b$ are equal [15], see also Appendix A, where we give the proof for completeness.

As a direct consequence of Eq. (8), partitioning H always increases the sampling error independently of the state $|\psi\rangle$ and partitioning $\mathcal{B}$ (see also Appendix A),

$$\frac{1}{M} \mathrm{Var}_{|\psi\rangle}(H) \leq \sum_{H_b \in \mathcal{B}} \frac{\mathrm{Var}_{|\psi\rangle}(H_b)}{M_b}. \tag{9}$$

Further, Eq. (8) motivates the definition of the following cost function to compare measurement strategies, being different splittings of H, that is not only independent of the specific measurement budget allocation M_b but also independent of the specific number of measurements M.

Definition 1 (Relative sampling complexity (pure state version)). Let $\mathcal{B}_1$ and $\mathcal{B}_2$ be partitionings of H. We call the ratio of the number of samples $M_{\mathcal{B}_1}/M_{\mathcal{B}_2}$ to obtain the same standard error $\sigma_{\mathcal{B}_1}^2 = \sigma_{\mathcal{B}_2}^2$ [see Eq. (8)], the *relative sampling complexity*,

$$\mathcal{G}_{|\psi\rangle}(\mathcal{B}_1, \mathcal{B}_2) := \frac{M_{\mathcal{B}_1}}{M_{\mathcal{B}_2}} = \left[\frac{\sum_{H_b \in \mathcal{B}_1} \sqrt{\mathrm{Var}_{|\psi\rangle}(H_b)}}{\sum_{H_b' \in \mathcal{B}_2} \sqrt{\mathrm{Var}_{|\psi\rangle}(H_b')}} \right]^2. \tag{10}$$

Note that the second equality holds thanks to Eq. (8). For $\mathcal{G}_{|\psi\rangle}(\mathcal{B}_1, \mathcal{B}_2) > 1$, we thus get a sampling improvement for partitioning $\mathcal{B}_2$ compared to partitioning $\mathcal{B}_1$, and the value of $\mathcal{G}_{|\psi\rangle}(\mathcal{B}_1, \mathcal{B}_2)$ tells us how much less often we have to measure within partitioning $\mathcal{B}_2$ compared to $\mathcal{B}_1$ for the same standard error, see Eq. (4), under optimal sampling budget allocation. Our results can also be interpreted as quantifying which strategy leads to a smaller standard error for the same number of measurements M. Equally, one can view $\mathcal{G}_{|\psi\rangle}(\mathcal{B}_1, \mathcal{B}_2)$ as the relative efficiency of unbiased estimators $\overline{H}_M^{\mathcal{B}_1}$ and $\overline{H}_M^{\mathcal{B}_2}$, as it is also equal to $\sigma_{\mathcal{B}_1}^2/\sigma_{\mathcal{B}_2}^2$ for $M_{\mathcal{B}_1} = M_{\mathcal{B}_2} = M$.

III. SAMPLING IMPROVEMENTS FOR EXACT EIGENSTATES

Many quantum algorithms aim to prepare eigenstates of Hamiltonians, including ground states. In this section, we thus consider the task of estimating energies E_i

of eigenstates, where $H|E_i\rangle = E_i|E_i\rangle$. Our aim is to make statements about the sampling improvement factor $\mathcal{G}_{|E_i\rangle}\left(\mathcal{B}_{\text{Pauli}}, \mathcal{B}_{L_x,L_y}\right)$, where $\mathcal{B}_{\text{Pauli}}$ labels the partitioning of H into fewest groups of mutually commuting local operators, which often coincide for lattice Hamiltonians with partitioning into different Pauli operators. $\mathcal{B}_{L_x,L_y}$ labels geometric partitioning, where each partition is obtained by splitting of H into disjoint patches of size $L_x \times L_y$, compare Fig. 1. For explicit definitions see Appendix B 1. For all considered partitionings of H, we have the following lemma.

Lemma 1 (Variance (in-)equalities). For a state vector $|\psi\rangle$ and Hamiltonian partitioning $H = \sum_{H_b\in\mathcal{B}} H_b$ we have

$$|\mathcal{B}| \sum_{H_b\in\mathcal{B}} \mathrm{Var}_{|\psi\rangle}(H_b) \geq \left(\sum_{H_b\in\mathcal{B}} \sqrt{\mathrm{Var}_{|\psi\rangle}(H_b)}\right)^2 \tag{11}$$

$$\geq \sum_{H_b\in\mathcal{B}} (2b-1)\,\mathrm{Var}_{|\psi\rangle}(H_b), \tag{12}$$

where $b \in \{1, 2, \ldots, |\mathcal{B}|\}$. The variance terms are, without loss of generality, taken to be ordered nonincreasingly as $\mathrm{Var}_{|\psi\rangle}(H_1) \geq \mathrm{Var}_{|\psi\rangle}(H_2) \geq \ldots$.

Proof. Equation (12) is a purely algebraic observation. Let us denote $\mathrm{Var}_{|\psi\rangle}(H_b) := a_b \geq a_{b+1} \geq 0$ and $|\mathcal{B}| = B$. From $0 \leq (a_b - a_k)^2 = (a_b + a_k)^2 - 4a_b a_k$ follows $\sqrt{a_b a_k} \leq \frac{1}{2}(a_b + a_k)$ and hence

$$\left(\sum_{b=1}^{B} \sqrt{a_b}\right)^2 \leq \sum_{b=1}^{B} a_b + \sum_{b>k=1}^{B} (a_b + a_k) = B\sum_{b=1}^{B} a_b.$$

Similarly, from $\sqrt{a_b a_k} \geq \min(a_b, a_k)$ follows

$$\left(\sum_{b=1}^{B} \sqrt{a_b}\right)^2 \geq \sum_{b=1}^{B} a_b + \sum_{b=2}^{B} 2(b-1)a_b = \sum_{b=1}^{B} (2b-1)a_b.$$

■

As an intermediate consequence for any partitioning scheme into two parts, and energy estimation of an eigenstate, we know that we optimally have to measure both parts equally often, as both variances are the same.

Corollary 1 (Variance equality for eigenstates and two partitionings). Let H be partitioned into two parts H_1 and H_2 and $|\psi\rangle = |E_i\rangle$ be an eigenstate of H, then

$$\mathrm{Var}_{|E_i\rangle}(H_1) = \mathrm{Var}_{|E_i\rangle}(H_2). \tag{13}$$

Proof. For bipartitions of eigenstates we obtain from $\mathrm{Var}_{|E_i\rangle}(H) = 0$,

$$\mathrm{Var}_{|E_i\rangle}(H_1) + \mathrm{Var}_{|E_i\rangle}(H_2) = 2|\mathrm{CoV}_{|E_i\rangle}(H_1, H_2)|$$
$$\leq 2\sqrt{\mathrm{Var}_{|E_i\rangle}(H_1)\,\mathrm{Var}_{|E_i\rangle}(H_2)}, \tag{14}$$

where we used that the covariance is an inner product and applied the Cauchy-Schwarz inequality. From $\sqrt{ab} \leq \frac{1}{2}(a+b)$, Eq. (11), we obtain that Eq. (14) needs to be an equality, and hence Eq. (13). ■

We note that Corollary 1 also implies that the commutator-based lower bound (commonly named uncertainty principle [24]) vanishes since $\mathrm{CoV}_{|E_i\rangle}(H_1, H_2) = \mathrm{CoV}_{|E_i\rangle}(H_2, H_1)$, and thus $|\langle E_i|\,[H_1, H_2]\,|E_i\rangle| = 0$.

Furthermore, we emphasize that even if a Pauli partitioning $\mathcal{B}_{\text{Pauli}}$ requires more than two noncommuting Pauli groupings, we can still restrict our considerations to the case $H = H_1 + H_2$. To see this, let us assume we need to partition the mutually commuting grouping measurements into three parts. As Eq. (9) is independent of the state, we know that measuring finer partitions always increases the combined variance (e.g., insert H_2 instead of H on the l.h.s.). Hence, for finding the lower bound, it suffices to consider a two-partition, where we can measure H_1 and put all other terms into H_2.

We consider geometric partitionings that split H into two equal parts, $H = H_1 + H_2$ with $|\mathcal{B}| = 2$,

$$H_1 := \frac{1}{2}\left(H - H_{\text{cut}} + H'_{\text{cut}}\right), \tag{15}$$

$$H_2 := \frac{1}{2}\left(H + H_{\text{cut}} - H'_{\text{cut}}\right), \tag{16}$$

where H_{cut} and H'_{cut} are sums of the 2-local terms in H, which are cut into their local parts in H_1 and H_2, respectively. In particular, we consider cutting the lattice into one-dimensional slices of thickness L, two-dimensional patches of size $L_x \times L_y$ and 2-local dimers of lattice sites. As an example, a partitioning into slices of thickness $L \geq 2$ is explicitly given by

$$H_{\text{cut}} = \sum_{m=1}^{\lfloor n_x/L \rfloor} \sum_{\ell=1}^{n_y} \mathcal{H}_{(mL,\ell),(mL+1,\ell)}, \tag{17}$$

$$H'_{\text{cut}} = \sum_{m=1}^{\lfloor n_x/L \rfloor} \sum_{\ell=1}^{n_y} \mathcal{H}_{(mL-1,\ell),(mL,\ell)}, \tag{18}$$

with the $\mathcal{H}_{(i_x,i_y),(j_x,j_y)}$ as introduced in Eq. (5). We provide detailed descriptions of all considered partitionings in Appendix B 1 and Appendix D 3. Furthermore, the

correlation

$$\mathrm{CoR}_{|E_i\rangle}(H_{\mathrm{cut}}, H'_{\mathrm{cut}}) = \frac{\mathrm{CoV}_{|E_i\rangle}(H_{\mathrm{cut}}, H'_{\mathrm{cut}})}{\sqrt{\mathrm{Var}_{|E_i\rangle}(H_{\mathrm{cut}})\,\mathrm{Var}_{|E_i\rangle}(H'_{\mathrm{cut}})}} \quad (19)$$

will play a significant role in the bounds we derive below.

With these definitions, we can now quantify the ratios of required measurement numbers, $\mathcal{G}_{|\psi\rangle}(\mathcal{B}_1, \mathcal{B}_2)$, for the measurement strategies discussed in Fig. 1. We prove the following lower bounds on the relative sampling complexity (Definition 1) improvement.

Theorem 1 (Relative sampling complexity improvement lower bound). Let H be a translation-invariant 2-local Hamiltonian on an $n_x \times n_y$ rectangular lattice as specified by Eq. (5). Let $\mathcal{B}_{\mathrm{Pauli}}$ be a Pauli partitioning and $\mathcal{B}_{L_x,L_y}$ be a geometric partitioning. Then, for any nondegenerate eigenstate $|E_i\rangle$ of H,

$$\mathcal{G}_{|E_i\rangle}\left(\mathcal{B}_{\mathrm{Pauli}}, \mathcal{B}_{L_x,L_y}\right) \geq \begin{cases} 4L\dfrac{1}{1-\mathrm{CoR}_{|E_i\rangle}(H_{\mathrm{cut}}, H'_{\mathrm{cut}})}, & \text{(20a)} \\ 4\dfrac{L_xL_y}{L_x+L_y}\dfrac{1}{1-\mathrm{CoR}_{|E_i\rangle}(H_{\mathrm{cut}}, H'_{\mathrm{cut}})}, & \text{(20b)} \\ \dfrac{4}{3}, & \text{(20c)} \end{cases}$$

where the cases denote different types of geometric partitionings. (a) corresponds to cutting H in one direction so that $(L_x, L_y) = (n_x, L)$ or equivalently $(L_x, L_y) = (L, n_y)$ with L being the thickness of the 1D cuts. (b) corresponds to cutting H into rectangular patches of size $L_x \times L_y$ with $n_x/2 \geq L_x \geq 2$ and $n_y/2 \geq L_y \geq 2$. (c) corresponds to the equal splitting of H into four 2-local partitions, $(L_x, L_y) =$ 2-local. Further, the correlation is always non-negative, $\mathrm{CoR}_{|E_i\rangle}(H_{\mathrm{cut}}, H'_{\mathrm{cut}}) \geq 0$, and thus $\frac{1}{1-\mathrm{CoR}_{|E_i\rangle}(H_{\mathrm{cut}}, H'_{\mathrm{cut}})} \geq 1$.

Proof (sketch). In all three cases, the proof is a composition of the variance of the Hamiltonians in $\mathcal{B}_{\mathrm{Pauli}}$ and $\mathcal{B}_{L_x,L_y}$ into the ones of the unit cell interaction V. The multilinearity of the covariance and the invariance of the ground state enable an identification of terms yielding the respective factor for each case. See Appendix B 1 for technical details. For explicit examples of Pauli and geometric partitionings see Appendix D 3. ■

We have formulated Theorem 1 as a sample improvement when measuring the Hamiltonian on nondegenerate eigenstates. Note that case (a) is not a special case of (b), as the latter cuts in both directions, s.t. for example, setting $L_x = n_x$ will yield a weaker bound. It readily generalizes to other measurement schemes $\mathcal{B}$. A generalization from 2-local to k-local interactions is straightforward, albeit it comes with a weakening factor on the sample improvement, or respectively a scaling factor of necessary cluster thicknesses.

The lower bounds (20a) and (20b) in Theorem 1 become arbitrarily large for $\mathrm{CoR}(H_{\mathrm{cut}}, H'_{\mathrm{cut}}) \to 1$, or equivalently $\mathrm{Var}_{|E_i\rangle}(H^{(1)}_{L_x,L_y}) = \mathrm{Var}_{|E_i\rangle}(H^{(2)}_{L_x,L_y}) \to 0$. This behavior is expected for our strategy of comparing the required measurements to achieve the same precision. For $\sigma_{\mathcal{B}_{L_x,L_y}} \to 0$ one would need a diverging number of measurements via Pauli partitioning with $\sigma_{\mathcal{B}_{\mathrm{Pauli}}} \neq 0$ to achieve the same standard error, cf. Eq. (8).

We conclude that for pure eigenstates, geometric partitioning will always yield a sampling advantage compared to standard Pauli partitioning, provided it can be efficiently implemented. The implementability of the measurement circuits U_{H_b} is mainly determined by the extra noise they generate due to imperfect implementations of the involved gates. To answer this question and analyze the impact of errors in the preparation of the measured state, we will cover in the following how the above results generalize to imperfect state preparation and imperfect implementation of U_{H_b}.

Next, we provide numerical examples illustrating our analytical results. Readers who are primarily interested in the generalization of our analytical results to scenarios that take imperfections in the state preparation and readout circuits into account may directly continue with Sec. IV. All examples considered in this section are on a two-dimensional $(n_x, n_y) = (4, 6)$ rectangular lattice with periodic boundaries, apart from the (spinful) Fermi-Hubbard model, which also uses 24 qubits on a 3×4 lattice. The considered examples are the 2D transverse field XY model (Sec. III A), the 2D transverse field Ising model (Sec. III B), and the 2D transverse field biaxial next-nearest neighbor Ising model (Sec. III C), which are illustrated in Fig. 2. In Fig. 4, we show the hard-core Bose-Hubbard model (Sec. III D), the spinless (Sec. III E) and the spinful Fermi-Hubbard model (Sec. III F).

A. Two-dimensional transverse field XY model

Theorem 1 indicates that measuring via geometric partitionings can become (almost) as good as measuring in the eigenbasis. The relative sampling complexity improvement lower bound comparing geometric partitioning with mutually commuting local operator grouping, which mostly coincides with Pauli partitioning for simple lattice models, can even diverge for $\mathrm{CoR}(H_{\mathrm{cut}}, H'_{\mathrm{cut}}) \to 1$. This means that a specific eigenstate correlates parts of the Hamiltonian H_{cut} and H'_{cut} very efficiently. Plausible candidates are highly entangled states, for instance, around zero-temperature (= ground state) phase transitions, where band gaps typically close. If H_{cut} and H'_{cut} have at least one lattice site in common, then their graph distance is 0. This applies, for example, for $L = 1$ or $(L_x, L_y) = (2, 2)$.

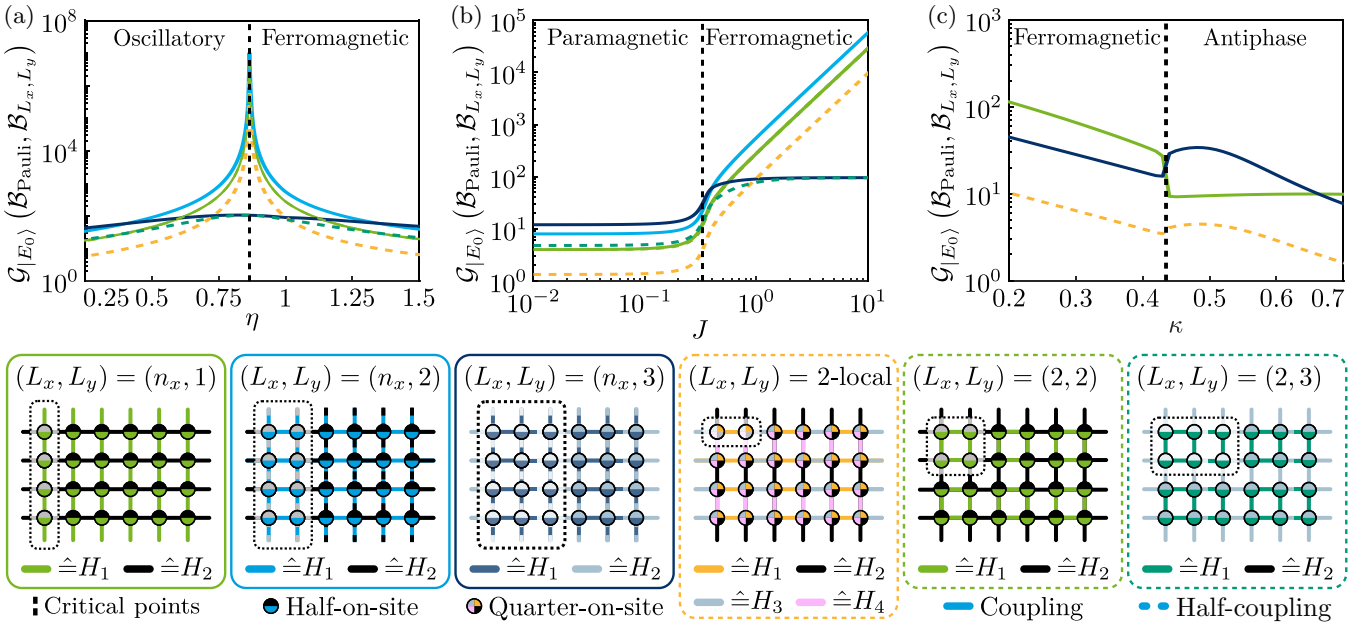

FIG. 2. Numerical examples: transverse field XY model, transverse field Ising model, and transverse field biaxial next-nearest-neighbor Ising. These are examples for the application of Theorem 1. All examples are on a 4×6 rectangular lattice with periodic boundary conditions. Critical points of phase transitions are indicated with dashed vertical lines and the different geometric splittings are color coded (compare graphical legend as well as Tables I–III in Appendix D 3). In the legend, the dotted boxes show the size of the unit cell, for which one needs to find a U_m in a geometric partitioning, compare also Fig. 1. Lattice vertices that are colored with two colors indicate that the geometric partitioning splits the single-site terms symmetrically into two parts. For 2-local partitionings, the single-site terms are symmetrically split into four parts. (a) 2D transverse field XY model: For $\eta \to \eta_c$ (here: $\eta_c = \sqrt{3}/2$) the improvement increases and diverges around the phase transition, where correlation lengths increase (see Theorem 1). (b) 2D transverse field Ising model. For $J/h \ll 1$ this is an example where lower bounds are approximately realized. In contrast, for $J/h \gg 1$, we see improvements noticeably surpassing the lower bounds, with scalings of $16L$ up to $\mathcal{O}((J/h)^2)$. For details see Lemma 2. Here, the Pauli cost is multiplied by 100 to fit the common scale. (c) 2D transverse field biaxial next-nearest-neighbor Ising model. κ is the next-nearest-neighbor coupling, where $\kappa = 0$ corresponds to the transverse field Ising model. As the interaction range for this model is that of a 3-local Hamiltonian, fewer geometric partitionings are available for the considered lattice sizes.

Then, e.g., the Hastings-Koma Lemma [25] does not yield an upper bound on the covariance making these specific geometric partitionings particularly well-suited candidates for divergent relative sampling complexity improvements.

Indeed, the 2D transverse field XY model [26], provides such an example, which is described by the following Hamiltonian:

$$H_{\text{TFXYM}} = -\frac{1}{2}\sum_{\langle i,j\rangle}\left((1+\eta)X_iX_j + (1-\eta)Y_iY_j\right) - h\sum_i Z_i, \tag{21}$$

with anisotropy parameter η. Note that for $\eta = 1$, we would obtain a transverse field Ising model with $J = h = 1$, which we show as a second example (see Sec. III B). H_{TFXYM} has three phases: a paramagnetic, a ferromagnetic, and an oscillatory one. This "oscillatory" phase is a type of ferromagnetic phase with oscillating band gap [27]. The phase boundary between ferromagnetic and oscillatory phase is at $\eta^2 + (h/2)^2 = 1$ [26].

Hence, we fix η or h to scan the other parameter over the phase transition. In Fig. 2(a) we set $h = 1$ and scan η. Indeed, we find around the phase transition, $\eta \approx \sqrt{3}/2$, as predicted by Theorem 1 for $L = 1$ (light green), $L = 2$ (cyan) as well as $(L_x, L_y) = (2, 2)$ (dashed light green, nearly not distinguishable from $L = 1$) a divergence of the relative sampling improvement compared to Pauli partitioning. In addition, this divergence is also realized for the geometric 2-local partitioning (dashed yellow), providing an example that also the easier-to-implement (c)-type measurement strategies in Theorem 1 can exhibit diverging relative sampling complexity improvements, even though we were not able to show this analytically. Further, the transverse field XY model (TFXYM) is an example for the extension of Theorem 1, which is proven for Pauli-2-partitions, onto $\geq$ 2-Pauli-partitions due to Eq. (9).

B. Two-dimensional transverse field Ising model

Another class of cases with large sampling improvement are examples where $\text{Var}_{|E_i\rangle}(H_1) = \text{Var}_{|E_i\rangle}(H_2)$ scale more beneficial than Pauli partitioning. One such instance

is a standard test bed of quantum spin models, the 2D transverse field Ising model [28–30],

$$H_{\mathrm{TFIM}} = -J \sum_{\langle i,j \rangle} Z_i Z_j - h \sum_i X_i, \tag{22}$$

which has a phase transition at $h_c \approx 3.044J$ [31], where it exhibits a ferromagnetic phase for $h < h_c$ and a paramagnetic phase for $h > h_c$. We show the following Lemma:

Lemma 2 (Perturbative improvement estimations). The sampling complexity improvement of estimating the ground state energy E_0 of the 2D transverse field Ising model (TFIM) on a rectangular lattice of geometric partitioning compared to Pauli sampling is given by

$$\mathcal{G}_{|E_0\rangle}\left(\mathcal{B}_{\mathrm{Pauli}}, \mathcal{B}_{L_x,L_y}\right) = \begin{cases} 4L + \mathcal{O}(J/h) & J/h \ll 1, \quad (23a) \\ \mathcal{O}((J/h)^2) & h/J \ll 1, L \leq 2, \quad (23b) \\ 32L + \mathcal{O}((h/J)^2) & h/J \ll 1, L > 2. \quad (23c) \end{cases}$$

Proof (sketch). The Lemma follows from tedious, but straightforward second perturbation theory calculations of the respective variances. See Appendix D. ■

This means that for $J/h \ll 1$ the lower bounds (a) in Theorem 1 are saturated, and hence tight. As shown in Fig. 2(b), in all three cases: 1D cuts, 2D cuts and 2-local partitions, the numerical results saturate the lower bounds, indicating that we cannot hope for larger lower bounds within the given setting.

Furthermore, for $J/h \gg 1$ we find within the numerical example for $L \leq 2$, $(L_x, L_y) = (2, 2)$ as well as 2-local geometric partitioning a model parameter dependent improvement of order $\mathcal{O}((J/h)^2)$, which matches Lemma 2 for $L \leq 2$. Even though our Theorem 1 gives a stronger lower bound for larger geometric partitions, this is an example of the possibility that smaller and thus easier to implement geometric partitions can yield a higher relative sampling complexity advantage. The respective decompositions of the local measurement unitaries U_b are given in Fig. 3. The 2-local measurement can be decomposed into just 2 CZ-gates and some 1-qubit gates. The 2×2 partition, where Theorem 1 guarantees at least a 4-fold reduction of the measurement effort, is equal to a 4 qubit chain with periodic boundary conditions. For this case, a particular efficient decomposition requiring only seven 2-qubit unitaries is known [20].

C. Two-dimensional transverse field biaxial next-nearest-neighbor Ising model

To arrive at a model, which we need to partition similarly to an (axial) 3-local one, we can add an axial next-nearest-neighbor Z_iZ_j interaction to the TFIM. Then, we obtain the transverse field biaxial next-nearest-neighbor Ising (TF-BNNNI) model on a two-dimensional rectangular lattice [36,37],

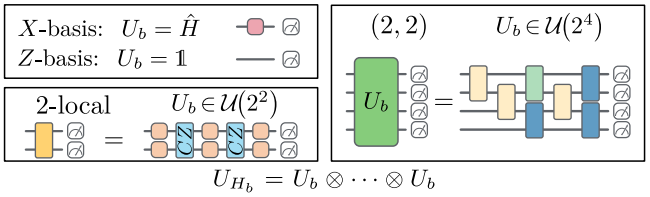

FIG. 3. Circuit decomposition example for the transverse field Ising model. Here, we give decompositions of the to-be-implemented local measurement unitaries U_b. For Pauli measurements, we can either just measure (Z) or need to apply Hadamard gates on each qubit (X). For 2-local geometric partitioning, we can decompose U_b into 1-qubit gates and 2 CZ-gates. For 4-qubit 2×2 unitaries, we would in general need to use exact [32,33] or approximate [22,34,35] compiling techniques. For the 2×2 transverse field Ising model, which is the same as a 4-qubit 1D transverse field Ising model with periodic boundaries, an efficient circuit using only seven 2-qubit unitaries is known [20].

$$H_{\text{TF-BNNNI}} = J\left(-\sum_{\langle i,j \rangle} Z_iZ_j + \kappa \sum_{\langle\langle i,j \rangle\rangle} Z_iZ_j\right) - h\sum_i X_i. \tag{24}$$

The TF-BNNNI model exhibits three distinct phases, compare, e.g., Fig. 2 of Ref. [37]. For $J = 1$ (to fix the energy scale), $h = 0$, and $\kappa < \kappa_c = \frac{1}{2}$, the nearest-neighbor coupling dominates and the ground state is ferromagnetically ordered. If κ becomes larger than 0.5, it becomes energetically favorable that 2×2 plaquettes are aligned and neighboring plaquettes are flipped with respect to each other. Thus, the BNNNI model can be viewed as a basic spin model for magnetic frustration.

Still, without a transverse field, the model remains diagonal or classical, and the eigenstate sampling task trivial as all eigenstates are computational basis states. In the limit $h \to \infty$ the ground state becomes a disordered X eigenstate. For our demonstration of how geometric partitioning works for a (quasi-)3-local model, we chose $h = 1$, and then $\kappa_c \approx 0.435$, compare Fig. 2(c). Compared to the 2-local next-nearest-neighbor models, we need to pick larger geometric partitions. The $L = 1$-partitioning (light green) still works as the NNN coupling is axial and both partitions are rotated with respect to each other. In contrast, the $L = 2$-cutting would not allow to include the NNN-terms, which is why the only other one-dimensional cut we consider is $L = 3$ (dark blue, leading to 3 parts). Similarly, the four 2-local partitions become six axial 3-local partitions

(dotted yellow) or for larger systems four axial 5-local partitions. For similar reasons, we do not show partitionings into two-dimensional patches, which would only be useful for larger lattices of 5×10 or 10×10 sites.

D. Two-dimensional hard-core Bose-Hubbard model

If we parameterize the TFXYM, see Eq. (21), with a symmetric interaction strength J,

$$H_{\mathrm{HCBH}} = -\frac{J}{2}\sum_{\langle i,j\rangle}\left(X_iX_j + Y_iY_j\right) + \frac{h}{2}\sum_i Z_i, \tag{25}$$

we obtain a Hamiltonian equivalent to the hard-core limit of the Bose-Hubbard model [38,39],

$$H_{\mathrm{HCBH}} = -J\sum_{\langle i,j\rangle}\left(a_i^\dagger a_j + a_i a_j^\dagger\right) + h\sum_i n_i, \tag{26}$$

with a_i, $a_i^\dagger$ and $n_i = a_i^\dagger a_i$ being bosonic annihilation, creation and number operators at lattice side i, respectively. $\langle i,j\rangle$ indicates nearest neighbors on a rectangular lattice. Compared to Bose-Hubbard model [40], in the hard-core limit the on-site repulsion term $U/2\sum_i n_i(n_i-1)$ is "frozen" out as $U\to\infty$. Hence, the system is described by Eq. (26) with the constraint that a lattice site can maximally be occupied by up to one boson due to the infinite on-site repulsion. This means that each lattice site is only a two-level system, and the Hilbert space of the model is only 2^n-dimensional. Further, the number operator $\sum_i n_i$ (or respectively $\sum_i Z_i$) commutes with the Hamiltonian, and thus gives an excitation number conservation symmetry, separating the Hilbert space into $\binom{n+1}{m}$-large sectors ($m \in \{0,..,n\}$) [41]. Hence, for $J \lessapprox h/4$, we are in the one-dimensional all (non)occupied sector, where the ground state is simply a Z eigenstate.

The numerical results for this model are shown in Fig. 4(a), where for $J \lessapprox h/4$ the ground state is a Z eigenstate. In turn, the configuration space becomes the largest, and hence the model most complex, if about half the sites are occupied ($m = n/2$), which is the case for the 4×6 hard-core Bose-Hubbard model for $0.4h \lessapprox J \lessapprox 0.47h$. Still, we find also in this most challenging regime sampling improvements of geometric partitioning $\mathcal{G}_{|E_i\rangle}\left(\mathcal{B}_{\mathrm{Pauli}}, \mathcal{B}_{L_x,L_y}\right)$ ranging from ca. 5 to 40.

E. Two-dimensional spinless Hubbard model

Next, we demonstrate that our geometric partitioning improves sampling efficiency for fermionic lattice models as well. For these models, we geometrically partition the Hamiltonian in the fermionic Fock basis and then perform a local fermion-to-qubit mapping for each patch separately (together with the transformations U_b). We start with the Hubbard model for spinless interacting fermions, that is similar to the hard-core Bose-Hubbard model from a lattice perspective, just that fermions live on the lattice sites,

$$H_{\text{Spinless Hubbard}} = -t\sum_{\langle i,j\rangle}\left(c_i^\dagger c_j + c_j^\dagger c_i\right) + U\sum_{\langle i,j\rangle} n_i n_j - \mu\sum_i n_i. \tag{27}$$

Here c_i, $c_i^\dagger$ and $n_i = c_i^\dagger c_i$ label annihilation, creation and number operators for fermions. The t-term implements fermionic hopping between neighboring sites on the rectangular lattice, where we set $t = 1$ to fix the energy scale. The U-term (for $U > 0$) penalizes the nearest-neighbor occupation [42]. Thus, for zero on-site energy $\mu = 0$, the model remains below half-filling ($\langle\sum_i n_i\rangle < n/2$). Similarly to the hard-core Boson Hubbard model, the number operator $\hat{N} = \sum_i n_i$ commutes with the Hamiltonian, is hence a symmetry, and as a consequence subdivides the Hilbert space into $\binom{n}{m}$, $m \in \{0,..,n\}$ large subspaces.

From this one would conclude that the largest sub-Hilbert space occurs for $m = n/2$, at half filling, but at this point the system has a further conserved quantity due to a particle-hole symmetry. Thus, we identify as a challenging scenario systems close to half-filling, but not exactly half-filled $n/2 \approx m \neq n/2$. For the 4×6 lattice model we consider here, this is realized for $\mu = 0$ and $U \lessapprox 0.6$, where we have $\langle\hat{N}\rangle = 11$ [compare Fig. 4(b)]. For this reason, we set in the following $\mu = 0$. Further, note that this does not make the (nongeometric) sampling problem easier as the number operators can be measured together anyway, and hence the number of standard partitions is unaffected, compare Appendix D 3. Similarly to the spirit of the baseline within the bosonic basis, we group here as a comparison the operators in $H_{\text{Spinless Hubbard}}$ in groups of mutually commuting operators. For $U \neq 0$ there are five groups, four groups being from the hopping part of the Hamiltonian with disjoint support, and one group measuring the number operators. Under Jordan-Wigner mapping, this corresponds to the Pauli grouping, which is why we still call it Pauli partitioning in Fig. 2.

Further, we find for $U \approx 1$ ($\langle\hat{N}\rangle = 8$), that the ground state of the model is degenerate or respectively gapless. Even though our Theorem 1 does then not strictly apply any longer, we find in this numerical test that the sampling improvement for geometric partitioning persists.

F. Two-dimensional spinful Hubbard model

As a final example for geometric partitioning to reduce the measurement effort when estimating energies of eigenstates, we show the Hubbard model. This model, which was originally proposed as an effective model to study electron correlations, for example how transitions from conductors to insulators occur [43], is of central

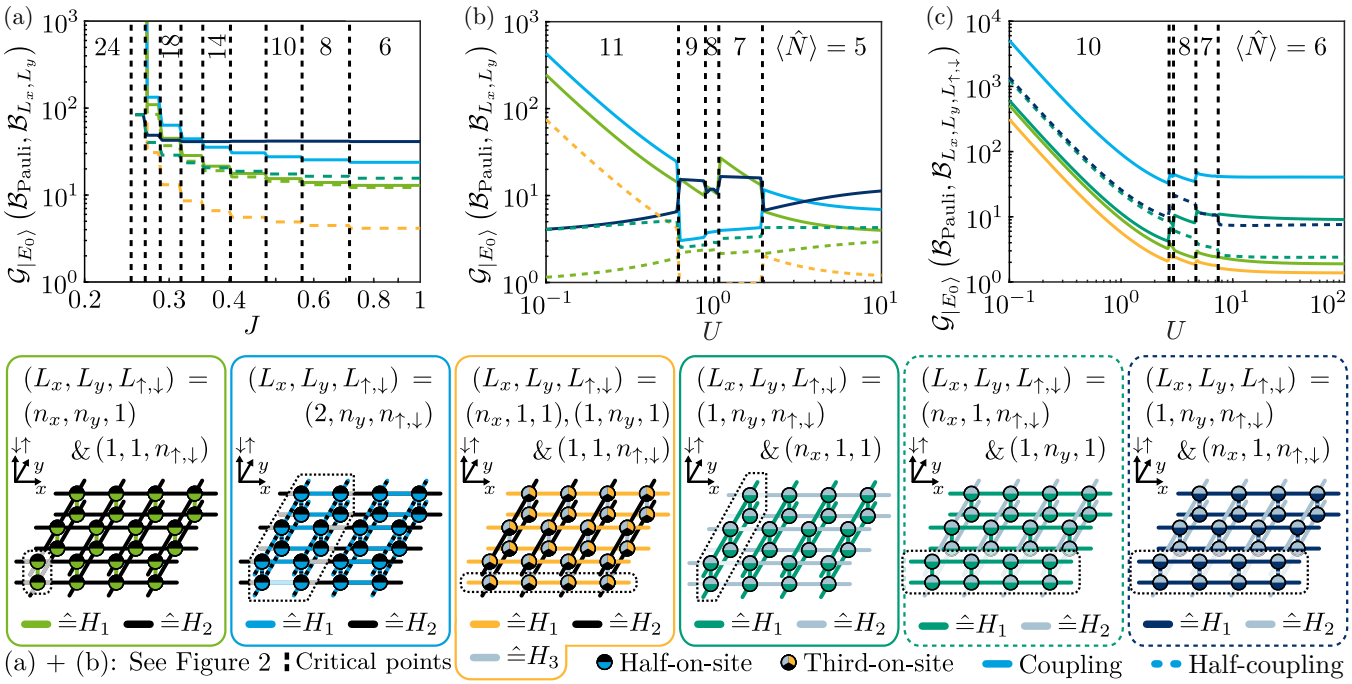

FIG. 4. Numerical examples: Hubbard-typed models. The spinless examples [(a) and (b)] are on a 4×6 rectangular lattice with periodic boundary conditions. Their partitionings and colors, including solid and dashed lines, correspond to the ones in Fig. 2. Dashed vertical lines indicate transitions between symmetry sectors of different particle number, and the different geometric splitting are color coded. (a) Hard-core Bose-Hubbard model. This model is similar to the transverse field XY model where the XY couplings are tuned symmetrically. The numbers show the amount of bosons in the system $\langle E_0|\hat{N}|E_0\rangle$. (b) Spinless Fermi-Hubbard model. This model resembles the hard-core Bose-Hubbard model, but with fermions instead of bosons on the lattice sides. Here, as a baseline the partitioning separating the Hamiltonian into mutually commuting groups of 2-local fermionic Hermitian operators is used. (c) (Spinful) Fermi-Hubbard model. In comparison with its spinless version, here the repulsive nearest-neighbor coupling is replaced with a repulsive on-site spin-spin coupling. Thus, it can be viewed as two spinless Hubbard models, coupled by their spin degree of freedom. Thus, it also consists of 24 lattice sides, but on a $4 \times 3 \times 2$ graph (compare graphical legend as well as Tables I, IV and V in Appendix D 3).

importance in condensed matter physics. It resembles the previously shown spinless fermion model, but features an on-site spin-spin (repulsive) interaction, instead of neighbor repulsion,

$$H_{\text{Hubbard}} = -t \sum_{\langle i,j\rangle,\sigma} \left(c^\dagger_{i,\sigma} c_{j,\sigma} + c^\dagger_{j,\sigma} c_{i,\sigma}\right) + U \sum_i n_{i,\uparrow} n_{i,\downarrow} - \mu \sum_{i,\sigma} n_{i,\sigma}. \tag{28}$$

One can view this (spinful) Hubbard model as two spinless ones (w.o. neighbor repulsion), that are repulsively coupled only at the same lattice sites i, forming a biplanar graph with the two two-dimensional planes being the up and down spin sectors. Thus, one can consider the Hubbard model as an example of geometric partitioning on a three-dimensional lattice. Testing a real three-dimensional example, including cubic patches, remains challenging as one would need four lattice sites in each dimensional direction or 64 overall.

We keep $t = 1$ and $\mu = 0$, and expect again to remain close to but below half-filling for all U. Indeed, for $U \lessapprox 2$ we have $\langle\hat{N}\rangle = 10$, which is still a larger Hilbert space than half-filling due to the particle-hole symmetry at half-filling that halves the dimension of the Hilbert subspace [44]. The baseline we compare geometric partitioning with works analogously to the spinless case, with four partitions to measure the hopping terms and one to measure the on-site repulsion.

For the spinful Hubbard model, all three particle sectors in the middle in Fig. 4(c), $8 \leq \langle\hat{N}\rangle \leq 10$, and $2.5 \lessapprox U \lessapprox 7.5$, are gapless. We see that the improvement noticeably leaves the smooth trajectory it has before and after, but nonetheless and despite these cases not being captured within Theorem 1, we find the sampling improvement by geometric partitioning persists.

IV. SAMPLING IMPROVEMENTS UNDER GLOBAL DEPOLARIZATION NOISE

In real-world sampling situations, where we aim to prepare eigenstates, we can typically not avoid small perturbations. These can originate from approximation errors in the state preparation $U_{|E_i\rangle}$ and measurement

circuits U_{H_b} as well as imperfect operations on the quantum hardware. While a detailed study of noise effects would require a hardware-specific noise model, the noise in local gates can often be well approximated by local depolarization noise [45,46]. Further, local depolarization converges already for $\Omega(\sqrt{n})$ deep 2D brick wall circuits to a global depolarization noise [46]. We thus consider a noise channel for global depolarization, which also provides an instance and hardware agnostic conservative estimate for approximation errors in the unitaries $U_{|E_i\rangle}$ and U_{H_b},

$$\Delta_\epsilon(|\psi\rangle\langle\psi|) = (1-\epsilon)\,|\psi\rangle\langle\psi| + \epsilon\frac{\mathbb{1}_d}{d}. \tag{29}$$

Δ_ϵ conserves the initial state with probability $(1-\epsilon)$, and maps to the maximally mixed state $\mathbb{1}_d/d$ with probability ϵ, which can be thought of as one of the 4^n-1 possible Pauli errors occurring with the same probability.

As it is a unital noise channel ($\Delta_{1-\epsilon}(\mathbb{1}_d/d) = \mathbb{1}_d/d$), we can model the results for the noisy states $\Delta_\epsilon(|\psi\rangle\langle\psi|)$ as averages over pure states,

$$|\tilde{\psi}(\epsilon)\rangle := \sqrt{1-\epsilon}\,|\psi\rangle + \sqrt{\epsilon}\,|\xi\rangle\,, \tag{30}$$

where the error states $|\xi\rangle$ are effectively Haar-random states. As we consider expectation values only up to second moments here, it suffices to demand that averages over the $|\xi\rangle$ equal averaging over the full Hilbert space up to second moments. More technically, this means that $|\xi\rangle$ is distributed according to a spherical 2-design in $\mathbb{C}^d$, compare also Appendix C 1.

If the main error contribution to the entire quantum circuit is due to the state preparation circuit $U_{|E_i\rangle}$, we have the same noise strength ϵ for both Pauli and geometric partitioning. We can however account for noise due to the additional gates, that geometric partitioning will use to rotate into the local eigenbases via U_{L_x,L_y}, by considering different noise levels for Pauli and geometric partitioning. We thus introduce the ratio $\gamma \in [0,1]$ of both noise levels,

$$\gamma := \frac{\epsilon_{\text{Pauli}}}{\epsilon_{L_x,L_y}}. \tag{31}$$

Assuming that the noise is dominated by gate errors due to imperfect implementation of the gates, we can, for each case, estimate the global depolarization noise strength ϵ via the average gate error ϵ_{gate} as well as the number of gates of the entire circuit. Since 1-qubit gate errors are typically one to two orders of magnitude smaller than 2-qubit gate errors in state-of-the-art hardware platforms [47], we only count 2-qubit gates and denote their number as $\mathcal{N}_{2q}$. We can thus estimate [46], $(1-\epsilon) \approx (1-\epsilon_{\text{gate}})^{\mathcal{N}_{2q}}$. Further, circuit approximation errors can be factored in by increasing the effective gate count. For instance, if we have a state preparation approximation error $\epsilon_{U_{|E_i\rangle}} = 2\%$ and a 2-qubit gate error $\epsilon_{\text{gate}} = 0.1\%$, we can take this into account by increasing $\mathcal{N}_{2q}(U_{|E_i\rangle})$ by 20. We can thus approximate γ by

$$\gamma \approx \frac{1-(1-\epsilon_{\text{gate}})^{\mathcal{N}_{2q}\left(U_{|E_i\rangle}\right)}}{1-(1-\epsilon_{\text{gate}})^{\mathcal{N}_{2q}\left(U_{L_x,L_y}\right)+\mathcal{N}_{2q}\left(U_{|E_i\rangle}\right)}}, \tag{32}$$

where $\mathcal{N}_{2q}(U)$ is the number of 2-qubit gates in the circuit U with $U_{|E_i\rangle}|0\rangle^{\otimes n} = |E_i\rangle$ being the preparation circuit preparing $|E_i\rangle$ from $|0\rangle^{\otimes n}$ and U_{L_x,L_y} the readout circuit for the geometric partitioning. As the rotation in local Pauli bases can be done with 1-qubit gates, we use $\mathcal{N}_{2q}(U_{\text{Pauli}}) = 0$. For $\mathcal{N}_{2q}\epsilon_{\text{gate}} \ll 1$ one can further approximate

$$\gamma \approx \frac{\mathcal{N}_{2q}(U_{|E_i\rangle})}{\mathcal{N}_{2q}(U_{|E_i\rangle}) + \mathcal{N}_{2q}\left(U_{L_x,L_y}\right)} \gtrapprox 1 - \frac{\mathcal{N}_{2q}\left(U_{L_x,L_y}\right)}{\mathcal{N}_{2q}(U_{|E_i\rangle})}. \tag{33}$$

$\gamma = 1/2$ would mean that readout via geometric partitioning uses as many 2-qubit gates as the state preparation circuit. Hence for small enough patches one typically has $\gamma > 1/2$. Our numerical results show that such patch sizes nonetheless lead to very significant sampling improvements.

To take depolarizing noise into account, we need to consider the statistical error average $\mathbb{E}$ over the corresponding ensemble of states. Nevertheless, a similar upper bound to Eq. (9) holds true:

$$\begin{aligned}\frac{1}{M}\mathbb{E}\left[\text{Var}_{|\tilde{\psi}\rangle}(H)\right] &\le \frac{1}{M}\mathbb{E}\left[\left(\sum_{H_b\in\mathcal{B}}\sqrt{\text{Var}_{|\tilde{\psi}\rangle}(H_b)}\right)^2\right] \\ &\le \frac{1}{M}\left[\sum_{H_b\in\mathcal{B}}\sqrt{\mathbb{E}\left[\text{Var}_{|\tilde{\psi}\rangle}(H_b)\right]}\right]^2 \\ &=: \overline{\sigma}^2_{\mathcal{B}},\end{aligned} \tag{34}$$

where we applied a second time the Cauchy-Schwarz inequality onto the expectation value over $|\tilde{\psi}\rangle$ to obtain the second inequality, which would be an equality if the different random variables $\sqrt{\text{Var}_{|\tilde{\psi}\rangle}(H_b)}$ were stochastically independent. See Appendix A for the explicit calculation. The above inequality allows us to expand the previously defined relative sampling complexity (Definition 1) from pure to mixed states:

Definition 2. [Relative sampling complexity (mixed state version)]. Let ϵ_1 and ϵ_2 be the global depolarization noise strength for measuring the quantum state $|\tilde{\psi}(\epsilon)\rangle$

[Eq. (30)] in the partitionings $\mathcal{B}_1$ and $\mathcal{B}_2$ of H. Then we define the ratio of the number of samples $M_{\mathcal{B}_1}/M_{\mathcal{B}_2}$ to obtain $\overline{\sigma}^2_{\mathcal{B}_1} = \overline{\sigma}^2_{\mathcal{B}_2}$, c.f. Eq. (34), as the relative sampling complexity,

$$\overline{\mathcal{G}}_{|\psi\rangle}\left(\mathcal{B}_1^{\epsilon_1}, \mathcal{B}_2^{\epsilon_2}\right) := \frac{M_{\mathcal{B}_1}}{M_{\mathcal{B}_2}} = \left[\frac{\sum_{H_b\in\mathcal{B}_1}\sqrt{\mathbb{E}\left[\mathrm{Var}_{|\tilde{\psi}(\epsilon_1)\rangle}(H_b)\right]}}{\sum_{H'_b\in\mathcal{B}_2}\sqrt{\mathbb{E}\left[\mathrm{Var}_{|\tilde{\psi}(\epsilon_2)\rangle}(H'_b)\right]}}\right]^2, \quad (35)$$

where the second equality holds because of Eq. (34).

One might wonder whether Eq. (35) is a reliable measure because $\overline{\sigma}_{\mathcal{B}}$ is only an upper bound to the average of variances, see Eq. (34). To confirm that $\overline{\sigma}_{\mathcal{B}}$ provides an accurate estimate for the error in an energy measurement by sampling from a quantum state, we compare it to the error of a simulated sampling process in Sec. IV B.

Note that $\overline{\mathcal{G}}_{|\psi\rangle}\left(\mathcal{B}_1^{\epsilon_1}, \mathcal{B}_2^{\epsilon_2}\right) \to \mathcal{G}_{|\psi\rangle}\left(\mathcal{B}_1, \mathcal{B}_2\right)$ for $\epsilon \to 0$. Moreover, even though we here focus on global depolarization noise, one could similarly study a relative sampling complexity with respect to two quantum channels Λ_1 and Λ_2 instead of the same channel Δ_ϵ with different noise strengths ϵ_1 and ϵ_2. The following Lemma allows us to resolve $\mathbb{E}\left[\mathrm{Var}_{|\tilde{\psi}\rangle}(H_b)\right]$:

Lemma 3 (Variance of isotropically perturbed states). Let $|\psi\rangle \in \mathbb{C}^d$ be some fixed reference state vector, $|\xi\rangle$ a random vector distributed according to a spherical 2-design in $\mathbb{C}^d$, O an observable, and $\epsilon \in [0,1]$. Then, for $|\tilde{\psi}\rangle := \sqrt{1-\epsilon}\,|\psi\rangle + \sqrt{\epsilon}\,|\xi\rangle$,

$$\mathbb{E}\left[\mathrm{Var}_{|\tilde{\psi}\rangle}(O)\right] = (1-\epsilon)\,\mathrm{Var}_{|\psi\rangle}(O) \quad (36a)$$
$$+ \epsilon(1-\epsilon)\langle\psi\,|O|\,\psi\rangle^2 \quad (36b)$$
$$+ \epsilon\|O\|_F^2/d + \mathcal{O}(1/d), \quad (36c)$$

where $\|\cdot\|_F/d$ denotes the Frobenius norm renormalized by a dimension factor $d = 2^n$ for system size n.

Proof (sketch). We use known results from representation theory, more concretely the Schur-Weyl duality, and the so-called swap-trick to compute the Hilbert space averages. For the explicit calculation see Appendix C 1. ■

Note that the setting in Lemma 3 is equivalent to the pure state $|\psi\rangle$ undergoing global depolarization noise of strength ϵ. Further, $\mathbb{E}\left[\mathrm{Var}_{|\tilde{\psi}\rangle}(O)\right] \geq \mathrm{Var}_{|\psi\rangle}(O)$, if $\mathrm{Var}_{|\psi\rangle}(O) \leq \|O\|_F^2/d + \mathcal{O}(1/d)$. This means that white noise increases the variance unless the noise-free state-specific sampling error is worse than state-independent average case sampling.

Unlike for perfect eigenstates, we have for $\epsilon \neq 0$ a nonzero shot-noise $\mathbb{E}\left[\mathrm{Var}_{|\tilde{E}_i\rangle}(H)\right] = \epsilon(1-\epsilon)E_i^2 + \epsilon\|H\|_F^2/d + \mathcal{O}(2^{-n}) \neq 0$, even for measurements in the eigenbasis of H. We can thus interpret Eq. (36a) as a contribution to $\mathbb{E}\left[\mathrm{Var}_{|\tilde{\psi}\rangle}(O)\right]$ coming from the quantum fluctuations of O in the state $|\psi\rangle$, whereas the contributions in Eqs. (36b) and (36c) are due to the noise. Indeed, in our further results, the following ratios between these contributions:

$$\epsilon_{I\leftrightarrow II} := \frac{4\,\mathrm{Var}_{|E_i\rangle}\left(H^{(1)}_{L_x,L_y}\right)}{E_i^2 + \|H\|_F^2/d}, \quad (37)$$

$$\epsilon_{II\leftrightarrow III} := \frac{4\,\mathrm{Var}_{|E_i\rangle}\left(H^{(1)}_{\mathrm{Pauli}}\right)}{E_i^2 + \|H\|_F^2/d}, \quad (38)$$

play a central role.

Thus, $\mathcal{B}_1 = \mathcal{B}_H$ provides a nontrivial finite upper bound for the sampling improvement [see also Fig. 5(c)]. This means that, in the case of $\epsilon \neq 0$, there will be both an upper and lower bound, compare also Appendix C.

Theorem 2 (Relative sampling complexity improvement under global depolarization noise). Let $\overline{\mathcal{G}}_{|E_i\rangle}\left(\mathcal{B}_1^{\epsilon_1}, \mathcal{B}_2^{\epsilon_2}\right)$ be the relative sampling complexity with noise strength ϵ_1 for partition $\mathcal{B}_1$, and ϵ_2 for $\mathcal{B}_2$ (Definition 2). Let the noisy state under global depolarization noise be given by $|\tilde{E}_i(\epsilon)\rangle = \sqrt{1-\epsilon}\,|E_i\rangle + \sqrt{\epsilon}\,|\xi\rangle$, with $|E_i\rangle$ an eigenstate of H. Then we have the following regimes:

(I.) For low randomness, $\epsilon \leq \epsilon_{I\leftrightarrow II}$ the sampling improvement is at worst a factor 2 less than for $\epsilon = 0$,

$$\overline{\mathcal{G}}_{|E_i\rangle}\left(\mathcal{B}^{\gamma\epsilon}_{\mathrm{Pauli}}, \mathcal{B}^{\epsilon}_{L_x,L_y}\right) \geq \frac{1}{2}\mathcal{G}_{|E_i\rangle}\left(\mathcal{B}_{\mathrm{Pauli}}, \mathcal{B}_{L_x,L_y}\right) \quad (39)$$

independently of γ, see Eq. (31).

(II.) For intermediate randomness, we have

$$\overline{\mathcal{G}}_{|E_i\rangle}\left(\mathcal{B}^{\epsilon}_{L_x,L_y}, \mathcal{B}^{\epsilon}_{H}\right) \leq 2 \Rightarrow \epsilon \geq \epsilon_{I\leftrightarrow II}, \quad (40)$$

$$\overline{\mathcal{G}}_{|E_i\rangle}\left(\mathcal{B}^{\gamma\epsilon}_{\mathrm{Pauli}}, \mathcal{B}^{\epsilon}_{H}\right) \geq 2 \Rightarrow \epsilon \leq \frac{1}{2-\gamma}\epsilon_{II\leftrightarrow III}. \quad (41)$$

(III.) For a randomness-dominated regime with $\epsilon \geq \epsilon_{II\leftrightarrow III}/(C-2\gamma)$, where $C > 2$ is a constant, the sampling efforts become dominated by noise, and

$$\overline{\mathcal{G}}_{|E_i\rangle}\left(\mathcal{B}^{\gamma\epsilon}_{\mathrm{Pauli}}, \mathcal{B}^{\epsilon}_{H}\right) \leq C, \quad (42)$$

for any bi-partition of H ($H = H_1 + H_2$), in particular also for $\mathcal{B}^{\epsilon}_{L_x,L_y}$.

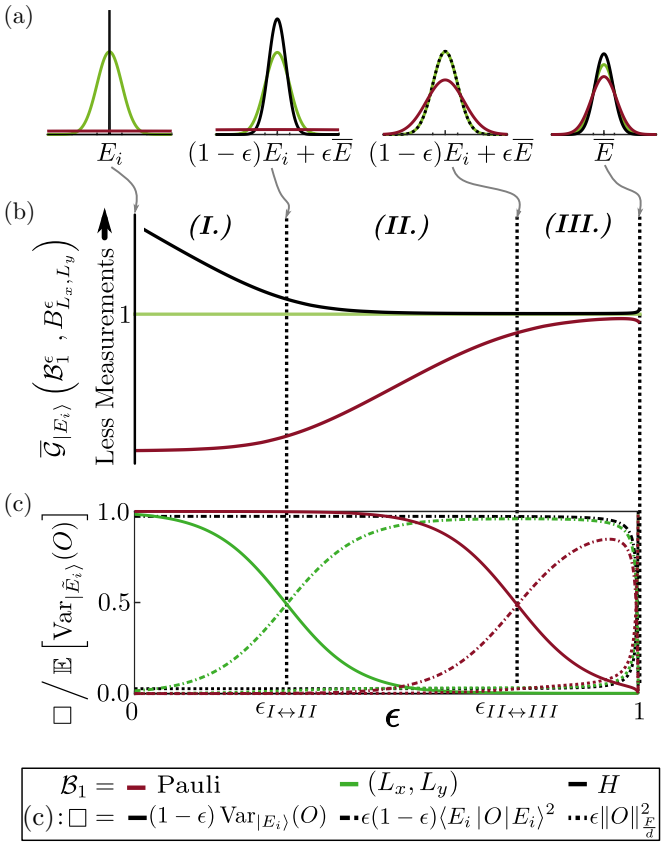

FIG. 5. Sampling improvement: from perfect eigenstates to Haar-random states: Here, we consider states of the form $|\tilde{E}_i\rangle = \sqrt{1-\epsilon}\,|E_i\rangle + \sqrt{\epsilon}\,|\xi\rangle$ with $|\xi\rangle$ Haar random up to second moments for a 2D transverse field Ising model. Yet all graphs will qualitatively look similar for different models. All three subplots share the same color coding: red for Pauli, green for geometric, and black for eigenbasis partitioning. Subplots (b) and (c) share the same x-axis (ϵ). In (a) we show the probability density corresponding to the empirical mean estimator $\overline{O}_M$ for different measurement strategies (color coded). The different snapshots depict (from left to right) $\epsilon = 0$, $\epsilon \approx \epsilon_{I\leftrightarrow II}$, $\epsilon \approx \epsilon_{II\leftrightarrow III}$ and $\epsilon \approx 1$. In (b) we provide a qualitative graph of the relative sampling complexity as defined in Definition 2. In agreement with Corollary 4, regime (II.) exists and has at least the size $\mathcal{G}_{|E_i\rangle}\left(\mathcal{B}_{\mathrm{Pauli}}, \mathcal{B}_{L_x,L_y}\right) \approx \epsilon_{II\leftrightarrow III}/\epsilon_{I\leftrightarrow II}$, if the noise mainly originates from imperfect state preparation, $\gamma \approx 1$, compare Eq. (32). In (c) we plot the relative weight of the terms in Eqs. (36a), (36b), and (36c) of Lemma 3, which are encoded by line type (solid, dashed-dotted and dotted), see legend. This shows that the transition from low to intermediate random regime (I.) → (II.) happens when the measurement uncertainty of geometric partitioning (green lines of the three types) becomes dominated by the noise-induced terms (36b) and (36c). Similarly, the transition from (II.) → (III.) corresponds to the same phenomena for Pauli partitioning (red lines). For reference, the noise-induced uncertainty of sampling in the eigenbasis is shown in black, which almost coincides with geometric partitioning in the middle of regime (II.) as well as regime (III.), as we expect from Theorem 2.

Proof (sketch). The proof in full detail is given in Appendix C 2. The basic idea is to use Lemma 3 to resolve $\mathbb{E}\left[\mathrm{Var}_{|\tilde{E}_i(\epsilon)\rangle}(H_b)\right]$ on the r.h.s. of Eq. (35). Then one uses Lemma 1 and Corollary 1 to linearize $(\sum \sqrt{.})^2$ both in the numerator and denominator, as well as

$$\left[\sum_k \sqrt{A_k + B_k}\right]^2 \geq \left[\sum_k \sqrt{A_k}\right]^2 + \left[\sum_k \sqrt{B_k}\right]^2. \quad (43)$$

To handle, the resulting fraction, we use iteratively the following algebraic inequality:

$$\frac{A}{A'} \geq \frac{B}{B'} \Rightarrow \frac{A}{A'} \geq \frac{A+B}{A'+B'} \geq \frac{B}{B'}, \quad (44)$$

for $A, A', B, B' \geq 0$. To see that Eq. (44) holds add AA' or BB' to $AB' \geq BA'$ and resolve for A/A' or B/B'. Eventually, to bound the resulting fractions, we use insights from our noise-free analysis, e.g., Theorem 1. ■

Note that regime (I.) does not depend on the noise level of Pauli partitioning $\gamma\epsilon$. Similarly, the high-noise threshold of regime (II.), $\propto 1/(2-\gamma)$, for $\gamma = 0$ is only decreased by a factor of 2 compared to $\gamma = 1$; if we were interested only in geometric partitioning performing better than Pauli partitioning even only a factor of 4/3. We have the following Corollary to determine whether geometric partitioning yields lower sampling noise than Pauli partitioning for any noise strength ϵ:

Corollary 2 (Criteria for sampling advantage of geometric partitioning under global depolarization). Let all objects be defined as in Theorem 2. Then we have

$$\overline{\mathcal{G}}_{|E_i\rangle}\left(\mathcal{B}_{\mathrm{Pauli}}^{\gamma\epsilon}, \mathcal{B}_{L_x,L_y}^{\epsilon}\right) \geq 1, \quad (45)$$

if

$$\gamma\left(\frac{\sum_{H_b\in\mathcal{B}_{\mathrm{Pauli}}}\|H_b\|_F}{\sum_{H_b'\in\mathcal{B}_{L_x,L_y}}\|H_b'\|_F}\right)^2 \geq 1, \quad (46)$$

$$\text{and} \quad \gamma(1+\alpha_{\mathrm{Pauli}}^2) \geq 1, \quad (47)$$

with $\alpha_{\mathrm{Pauli}} := \left|\langle H_{\mathrm{Pauli}}^{(1)}\rangle/E_i - 1/2\right| \geq 0$;
or if

$$\epsilon \leq c_\epsilon \epsilon_{II\leftrightarrow III}, \quad (48)$$

with $c_\epsilon = \max(\frac{1}{2-\gamma}, \left[1 - 1/\mathcal{G}_{|E_i\rangle}\left(\mathcal{B}_{\mathrm{Pauli}}, \mathcal{B}_{L_x,L_y}\right)\right])$.

Proof (sketch). Similarly to the methods for Theorem 2, we use Eq. (43) to linearized the noisy sampling error of

Pauli partition [Eq. (34)], and then Eq. (44) iteratively to obtain

$$\overline{\mathcal{G}}_{|\tilde{E}_i\rangle}\left(\mathcal{B}_{\text{Pauli}}, \mathcal{B}_{L_x,L_y}\right) \geq \min\left(\frac{A'}{A}, \frac{B'}{B}, \frac{C'}{C}\right), \tag{49}$$

with A, B, C according to the three lines in Lemma 3. $A'/A \geq \mathcal{G}_{|E_i\rangle}\left(\mathcal{B}_{\text{Pauli}}, \mathcal{B}_{L_x,L_y}\right)$ is larger than 1 due to Theorem 1, $B'/B \geq \gamma(1+\alpha^2_{\text{Pauli}})$ as geometric partitioning equally distributes the energy ($\alpha^2_{L_x,L_y} = 0$) and $C'/C \geq \gamma\left(\sum_{H_b\in\mathcal{B}_{\text{Pauli}}}\|H_b\|_F \Big/ \sum_{H'_b\in\mathcal{B}_{L_x,L_y}}\|H'_b\|_F\right)^2$. The first part of, Eq. (48) corresponds to Eq. (41) and the second part one obtains in a similar calculation using $\overline{\mathcal{G}}_{|E_i\rangle}\left(\mathcal{B}^{\gamma\epsilon}_{\text{Pauli}}, \mathcal{B}^{\epsilon}_{L_x,L_y}\right) \geq \overline{\mathcal{G}}_{|E_i\rangle}\left(\mathcal{B}^{0\epsilon}_{\text{Pauli}}, \mathcal{B}^{\epsilon}_{L_x,L_y}\right)$. The proof in full detail is given in Appendix C 3. ■

For $\gamma \approx 1$, Eq. (47) is typically fulfilled as Pauli partitioning in general does not distribute the energy equally. Then, it is sufficient to compute Eq. (46), which can be calculated classically at scale for sparse observables such as lattice Hamiltonians. As we know the Pauli sum representation of our sparse observables H_b, and we have for Paulistring P_i and Kronecker delta $\delta_{i,j}$: $\text{Tr}(P_iP_j)/d = \delta_{i,j}$, calculating the Frobenius norm boils down to summing the squared Pauli sum coefficients $\|H_b\|^2_F/d = \sum_{c_jP_j\in H_b} c_j^2$. Indeed, $\sum_{H_b\in\mathcal{B}}\|H_b\|_F/d$ gives precisely the state-independent average sampling error for partitioning H according to splitting $\mathcal{B}$, which we would have obtained if we had averaged over the Hilbert space, instead of deriving state-tailored sampling bounds [15]. Therefore, Eq. (46) tells us, for $\gamma \approx 1$ on average what improvement we can expect by geometric partitioning compared to Pauli partitioning for an arbitrary state. Thus, even if we cannot take advantage of any specific state structure, we can still improve our measurement scheme on the observable partitioning level by optimizing the partitioning with respect to Eq. (46).

Complementary to Corollary 2, one can require a targeted relative sampling improvement and ask how large the gate error ϵ_{gate} and the number of gates $\mathcal{N}_{2q}(U_{|E_i\rangle})$ and $\mathcal{N}_{2q}\left(U_{L_x,L_y}\right)$ can be to guarantee such an improvement. This is answered by the following corollary:

Corollary 3 (Gate budget for noisy geometric partitioning). Let all objects be defined as in Theorem 2 and prior. Let G be a desired relative sampling improvement. Then

$$\overline{\mathcal{G}}_{|E_i\rangle}\left(\mathcal{B}^{\gamma\epsilon}_{\text{Pauli}}, \mathcal{B}^{\epsilon}_{L_x,L_y}\right) \geq G, \tag{50}$$

provided that

$$\mathcal{N}_{2q}\left(U_{L_x,L_y}\right) \lessapprox \frac{\log\left(1-c^{(G)}_{\epsilon}\epsilon_{II\leftrightarrow III}\right)}{\log(1-\epsilon_{\text{gate}})} - \mathcal{N}_{2q}(U_{|E_i\rangle}), \tag{51}$$

with $c^{(G)}_{\epsilon} = \max\left(1/(2G-\gamma), \left[1 - G/\mathcal{G}_{|E_i\rangle}\left(\mathcal{B}_{\text{Pauli}}, \mathcal{B}_{L_x,L_y}\right)\right]\right)$.

Proof (sketch). The first c_ϵ follows by replacing "2" in Eq. (41) with "2G" and plugging in Eq. (32). $G \leq \mathcal{G}_{|E_i\rangle}\left(\mathcal{B}_{\text{Pauli}}, \mathcal{B}_{L_x,L_y}\right)/2$ ensures that we are still in regime (II.). The second c_ϵ similarly takes Eq. (48) and replaces "≥ 1" with "$\geq G$." For more details see Appendix C 4. ■

One can see that for very good gate fidelities, $\epsilon_{\text{gate}} \to 0$, the upper bound in Eq. (51) diverges. This is consistent with the intuition of having an arbitrary gate budget for measurement in this case. In turn, for very noisy circuits, $\epsilon_{\text{gate}} \to 1$, the bound would become negative, excluding, as expected, any further gates for the measurement circuit. One can see from Eq. (51) that, in order to have a positive bound for $\mathcal{N}_{2q}\left(U_{L_x,L_y}\right)$, one must have

$$\mathcal{N}_{2q}(U_{|E_i\rangle}) \lessapprox \frac{\log\left(1-c^{(G)}_{\epsilon}\epsilon_{II\leftrightarrow III}\right)}{\log(1-\epsilon_{\text{gate}})}. \tag{52}$$

State preparation circuits of larger size will not allow for a subsequent measurement circuit. For $G = 1$, we thus reobtain the intuitive expectation that the gate error ϵ_{gate} needs to be smaller than the threshold $\epsilon_{II\leftrightarrow III}$ for $\mathcal{N}_{2q}(U_{|E_i\rangle}) \gtrapprox 1$.

In Theorem 2, one can think of regime (I.) as effectively noise-free. In regime (II.), measuring in geometric partitioning is effectively as good (up to a factor 2) as the optimal strategy, measuring in the eigenbasis. Finally, regime (III.) is too noisy for state-tailored shot noise improvements since the main sampling error will be $\propto \epsilon(1-\epsilon)E_i^2 + \epsilon\|H\|^2_F/d$, and hence independent of the chosen measurement strategy.

The regimes of randomness predicted by Theorem 2 are illustrated in Fig. 5 for the example of a 2D transverse field Ising model. The data clearly show the transitions between the regimes as predicted by Theorem 2. To see that indeed all three regimes are realized, note that for $\epsilon \to 0$ all but the variance-type terms vanish, and respectively for $\epsilon \to 1$, all but the norm-type terms vanish. The following Corollary, lower bounds the size of regime (II.), and hence guarantees its existence:

Corollary 4 (Existence and intermediate noise regime). Let all objects be defined as in Theorem 2. Then there exist noise thresholds $\epsilon^{\text{th}}_{L_x,L_y}, \epsilon^{\text{th}}_{\text{Pauli}} > 0$ s.t.:

$$\forall\epsilon \geq \epsilon^{\text{th}}_{L_x,L_y}: \quad \overline{\mathcal{G}}_{|E_i\rangle}\left(\mathcal{B}^{\epsilon}_{L_x,L_y}, \mathcal{B}_H\right) \leq 2, \tag{53}$$

$$\forall\epsilon \leq \epsilon^{\text{th}}_{\text{Pauli}}: \quad \overline{\mathcal{G}}_{|E_i\rangle}\left(\mathcal{B}^{\gamma\epsilon}_{\text{Pauli}}, \mathcal{B}_H\right) \geq 2, \tag{54}$$

and the ratio of those noise thresholds is given by

$$\frac{\epsilon^{\text{th}}_{\text{Pauli}}}{\epsilon^{\text{th}}_{L_x,L_y}} \geq \left(\frac{1}{2-\gamma}\right)\mathcal{G}_{|E_i\rangle}\left(\mathcal{B}_{\text{Pauli}}, \mathcal{B}_{L_x,L_y}\right). \tag{55}$$

For all but 2-local partitioning, the r.h.s. of Eq. (55) ≥ 2 due to Theorem 1.

Proof. This is a direct consequence of Theorem 2, as have found there explicit approximate thresholds of the form $\epsilon_{I\leftrightarrow II} \geq \epsilon^{\text{th}}_{L_x,L_y}$ and $\epsilon^{\text{th}}_{\text{Pauli}} \geq 1/(2-\gamma)\epsilon_{II\leftrightarrow III}$, from which the statement follows. ■

In the limit of large depolarization noise ($\epsilon \to 1$), one thus recovers the same estimates as via averaging over Hilbert space, compare Appendix A. Our results show that sampling improvements can be more significant for small noise regimes, as the structure of the observable and its eigenstate can be exploited.

A. Sampling improvements under noise: Numerical analysis

Here, we give numerical examples of energy estimations of eigenstates under global depolarization noise. Our main objective is to explore the tightness of the analytical bounds, mostly Theorem 2. Compared to the noiseless examples in Secs. III A–III F, we thus have two more variable parameters in our numerical study, the noise strength ϵ and relative noise strength γ, see Eq. (31). Even though, in real-world scenarios, measuring in geometric partitioning will have lower noise than measuring in the eigenbasis due to shallower measurement circuits, we use here the same noise strength for both as a very conservative approximation to illustrate when geometric partitioning

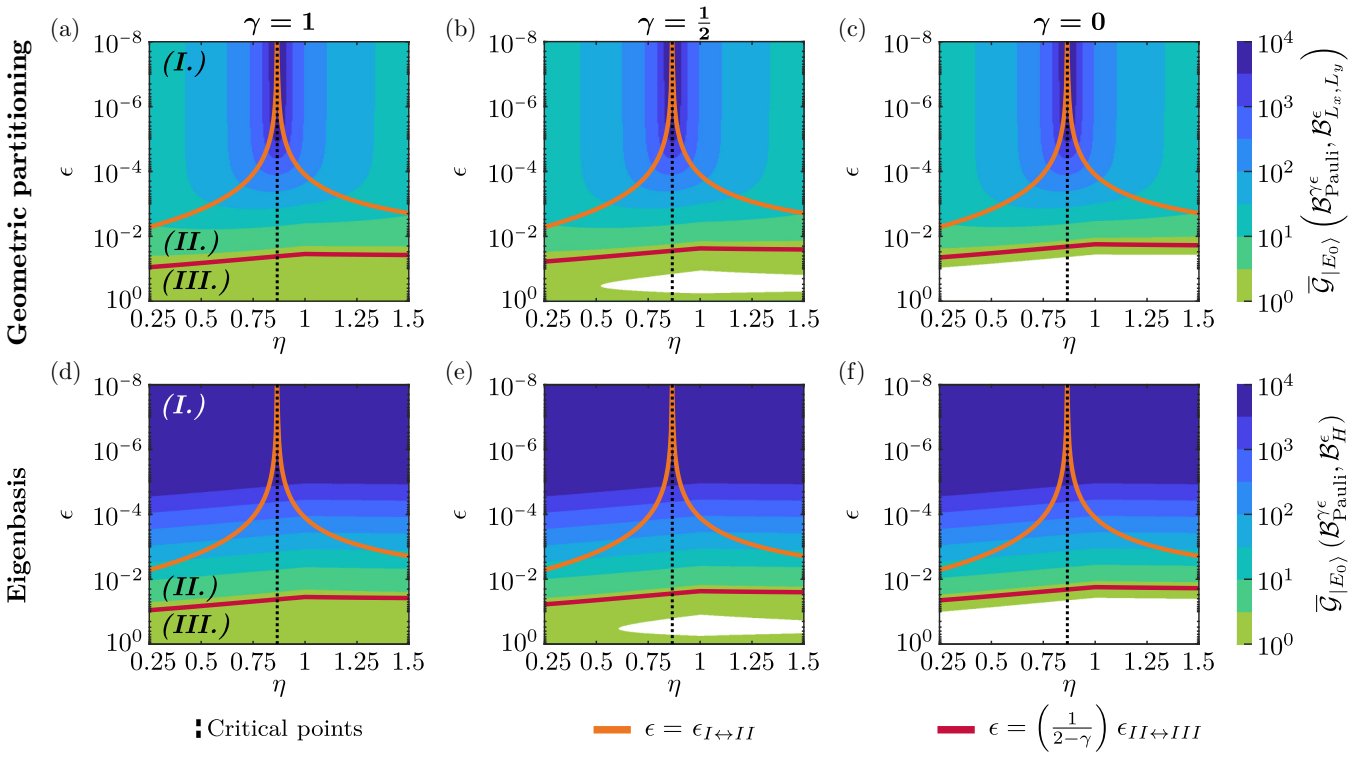

FIG. 6. Shot-noise reduction for the transverse field XY model under noise. Here we present numerical results that confirm Theorem 2 for the 2D XY model with Z field on a 4×6 rectangular lattice with periodic boundary conditions (see also Sec. III A and Fig. 2(a) for the $\epsilon = 0$ case). The vertical dashed lines mark, as in Fig. 2(a), the critical point at $\eta_c = \sqrt{3}/2$. We pick a geometric partitioning with $(L_x, L_y) = (2, 2)$, and plot relative sampling complexities $\overline{\mathcal{G}}_{|E_0\rangle}\left(\mathcal{B}^{\gamma\epsilon}_{\text{Pauli}}, \mathcal{B}^{\epsilon}_{L_x,L_y}\right)$ and $\overline{\mathcal{G}}_{|E_0\rangle}\left(\mathcal{B}^{\gamma\epsilon}_{\text{Pauli}}, \mathcal{B}^{\epsilon}_{H}\right)$, which are both color coded, for varying depolarization noise ϵ and system parameter η. The orange line marks $\epsilon = \epsilon_{I\leftrightarrow II}$ and the red line $\epsilon = \frac{1}{2-\gamma}\epsilon_{II\leftrightarrow III}$, see legend. We cut the improvement scale at 10^4 and a white area indicates that Pauli partitioning is better. The upper row shows the sampling improvement of geometric partitioning compared to Pauli partitioning, and the lower row compares hypothetical sampling in the eigenbasis with Pauli partitioning. The columns show different relative noise strength γ to take into account potentially additional gates when using geometric partitioning, or the eigenbasis. One can see that regime (I.) is effectively noiseless as there is no noticeable ϵ-dependence. Similarly, geometric partitioning (top row) is essentially as good as sampling in the true eigenbasis (bottom row) in regimes (II.) and (III.), as predicted by Theorem 2. In agreement with Corollary 2, we find that geometric partitioning is always better than geometric partitioning in the high-sampling advantage regimes (I.) and (II.). Moreover, the size of regime (II.) is given by $\frac{1}{2-\gamma}\mathcal{G}_{|E_0\rangle}\left(\mathcal{B}_{\text{Pauli}}, \mathcal{B}_{L_x,L_y}\right)$, as stated in Corollary 4. Therefore, a large noiseless sampling complexity improvement will lead to a large noise range where geometric partitioning performs effectively as well as measuring in the true eigenbasis and is guaranteed to be better than Pauli partitioning.

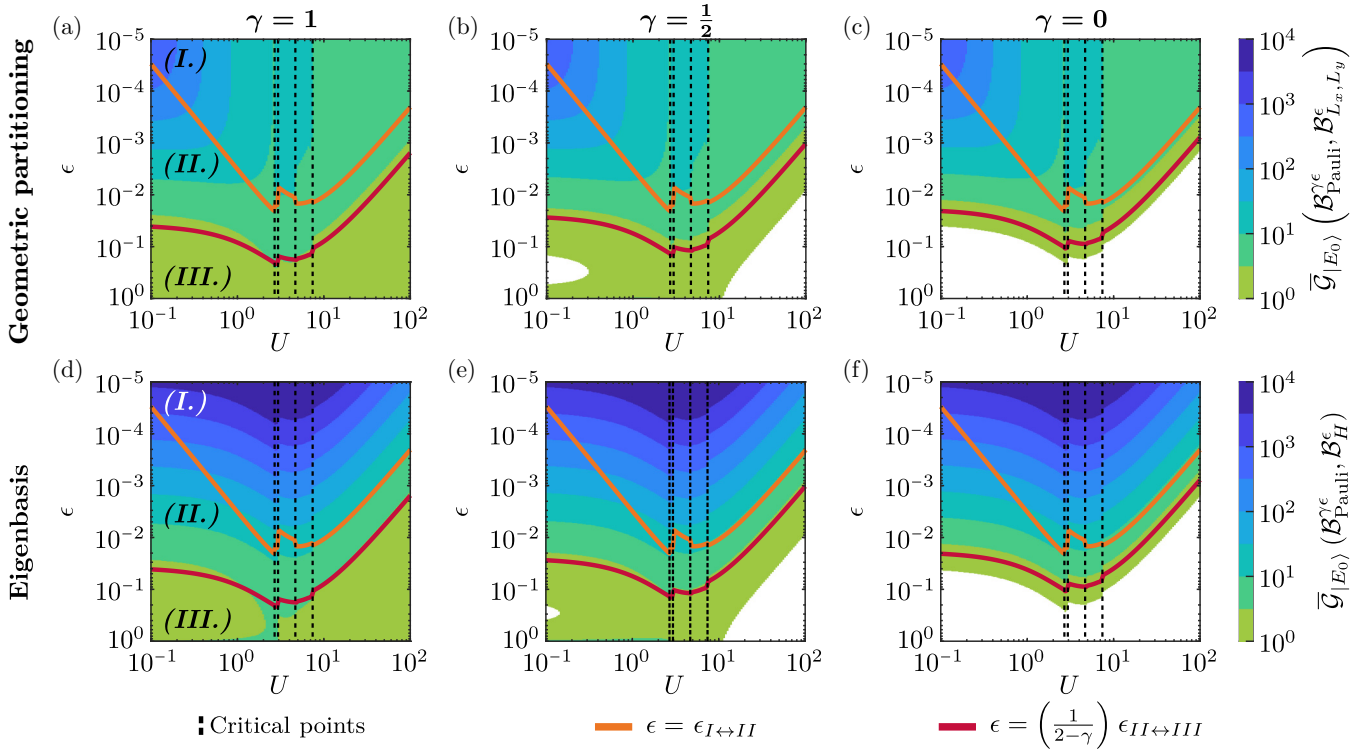

FIG. 7. Shot-noise reduction for the Fermi-Hubbard under noise. Here we show numerical results for the spinful Fermi-Hubbard model on a 3×4 lattice under noise, which corresponds to 24 qubits on a biplanar graph [The corresponding noise-free results are shown in Fig. 4(c)]. This figure is similarly structured as Fig. 6. Vertical dashed lines show, as in Fig. 4(c), the critical points, which are given by transitions of the average occupation number from $\langle \hat{N} \rangle = 10$ to $\langle \hat{N} \rangle = 6$ in integer steps. As the geometric partition we choose 1D cuts with $L = 1$ (which was seen to perform as well as $(L_x, L_y) = (2, 2)$ in all other examples) and plot relative sampling complexities $\overline{\mathcal{G}}_{|E_0\rangle}\left(\mathcal{B}^{\gamma\epsilon}_{\text{Pauli}}, \mathcal{B}^{\epsilon}_{L_x,L_y}\right)$ and $\overline{\mathcal{G}}_{|E_0\rangle}\left(\mathcal{B}^{\gamma\epsilon}_{\text{Pauli}}, \mathcal{B}^{\epsilon}_{H}\right)$ for varying depolarization noise ϵ and system parameter U. Similarly to the noisy transverse field XY model, we find that geometric partitioning yields lower shot noise for any ϵ for $\gamma = 1$, while $\gamma \neq 1$ only affects the high-noise regime (III.), where geometric partitioning already performs comparably to measuring in the eigenbasis. White areas correspond to regions where Pauli partitioning is better than geometric partitioning or eigenbasis measurements.

effectively works as well as the noisy eigenbasis. We focus our analysis here on one bosonic model, the TFXYM, see Fig. 6, and one fermionic model, the Fermi-Hubbard model, see Fig. 7.

In Figs. 6 and 7, all subfigures are structured in the same way. Along the horizontal axis, we vary a model parameter, the anisotropy η for the TFXYM and the interaction strength U for the Fermi-Hubbard model. Along the vertical axis, we vary the noise level in logarithmic scale. The color coding shows, in steps of half an order of magnitude, the relative sampling complexity $\overline{\mathcal{G}}_{|E_0\rangle}$ for the compared measurement strategies.

For exploring the transition between regimes (I.) and (II.) we compare three noise thresholds. The first is our approximation to the transition noise strength, $\epsilon_{I\leftrightarrow II}$ as defined in Eq. (37), the second is the noise level for which the sampling improvement is reduced by a factor two compared to the noiseless case and the third is the threshold above which geometric partitioning needs twice as many measurements as measuring in the eigenbasis. All three match very closely throughout most of the parameter ranges in Figs. 6 and 7, which is one of our main numerical findings, confirming that our analytical approximations are indeed tight. Thus, we draw in Figs. 6 and 7 only the line $\epsilon = \epsilon_{I\leftrightarrow II}$ in orange. The ϵ-values, where $\overline{\mathcal{G}}_{|E_0\rangle}\left(\mathcal{B}_{L_x,L^{\epsilon}_y}, \mathcal{B}^{\epsilon}_{H}\right) = 2$ and $\overline{\mathcal{G}}_{|E_0\rangle}\left(\mathcal{B}_{L_x,L^{\epsilon}_y}, \mathcal{B}^{\epsilon}_{H}\right) = \mathcal{G}_{|E_0\rangle}\left(\mathcal{B}_{\text{Pauli}}, \mathcal{B}_{L_x,L_y}\right)/2$, have relative deviations from the $\epsilon = \epsilon_{I\leftrightarrow II}$ values of at most 17% for the TFXYM, see Fig. 6, and at most 9.4% for $U \lessapprox 4.7$ for the Fermi-Hubbard model, see Fig. 7. For large U-values of the Fermi-Hubbard model (see Fig. 7) the ϵ-value for the threshold $\overline{\mathcal{G}}_{|E_0\rangle}\left(\mathcal{B}^{\gamma\epsilon}_{\text{Pauli}}, \mathcal{B}^{\epsilon}_{L_x,L_y}\right) = \mathcal{G}_{|E_0\rangle}\left(\mathcal{B}_{\text{Pauli}}, \mathcal{B}_{L_x,L_y}\right)/2$ is shifted toward larger ϵ by at most a factor 2.4 for $\gamma = 1$ and by a factor 1.5 for $\gamma = 1/2$, which is consistent with Theorem 2.

In regime (I.) we further expect the shot-noise reductions to differ at most by a factor two from the improvements in the noiseless case. Consequently, we expect $\overline{\mathcal{G}}_{|E_0\rangle}\left(\mathcal{B}^{\gamma\epsilon}_{\text{Pauli}}, \mathcal{B}^{\epsilon}_{L_x,L_y}\right)$ to barely change with ϵ and to

be independent of the relative noise strength of Pauli partitioning γ in this regime. This is confirmed by Figs. 6(a)–6(c) and Figs. 7(a)–7(c). For regimes (II.) and (III.), we expect that geometric partition performs up to a factor of two as good as measuring in the eigenbasis of H. Indeed, when we compare plots (a)–(d), (b)–(e), and (c)–(f) of Figs. 6 and 7, these regions appear very comparable. In agreement with Theorem 2, we only find for the high-noise regime (III.) and lower relative noise ($\gamma = 1/2$ and $\gamma = 0$), that Pauli partitioning becomes better than geometric partitioning (white areas). Yet here, measuring via Pauli splitting is typically also better than measuring in the eigenbasis.

In real-world settings and for small enough patch sizes, we expect the majority of the gates to be used for state preparation, and hence $\gamma \approx 1$. In such cases, we expect geometric partitioning to perform better than Pauli partitioning for any noise level ϵ due to Corollary 2. This can nicely be observed in Figs. 6(a) and 7(a). These results are in agreement with the observation that Eqs. (46) and (47) are fulfilled, as geometric splitting distributes the energy equally among the parts and also spreads the Pauli weights equally, so that typically also $\sum \|H_b\|_F$ is smaller for geometric than Pauli partitioning.

The case with $\gamma = 1/2$ corresponds to the same effective number of 2-qubit gates for both state preparation and geometric partitioning as shown in Eq. (32). We also show $\gamma = 0$, which compares noise-free perfect state preparation and Pauli measurement against noisy geometric partitioning not as a realistic application scenario, but to emphasize the validity of our analysis. Another observation in both noisy numerical examples is, that for $U/t, \eta = \mathcal{O}(1)$, the transition (II.) $\leftrightarrow$ (III.) typically lies at (large) noise strength of $\epsilon \approx 2\%$–10%. At the same time, the strongly correlated and thus particularly interesting regimes are typically also located in these parameter regimes.

When comparing the noisy sampling improvements here with their noiseless counterparts in Figs. 2(a) and 2(c), one can see, particularly for the TFXYM, a close relation between noisy and noiseless results as the large improvements carry over to the imperfect setting. Moreover, the size of regime (II.) is approximately equal to the value of $\mathcal{G}_{|E_0\rangle}\left(\mathcal{B}_{\text{Pauli}}, \mathcal{B}_{L_x,L_y}\right)$, as stated in Corollary 4. This means that for higher improvement of geometric partitioning in the noiseless case, is, the smaller the noise level ϵ needs to be for geometric partitioning performing recognizable worse than measuring in the eigenbasis. Note that this is independent of the relative noise strength of Pauli partitioning γ.

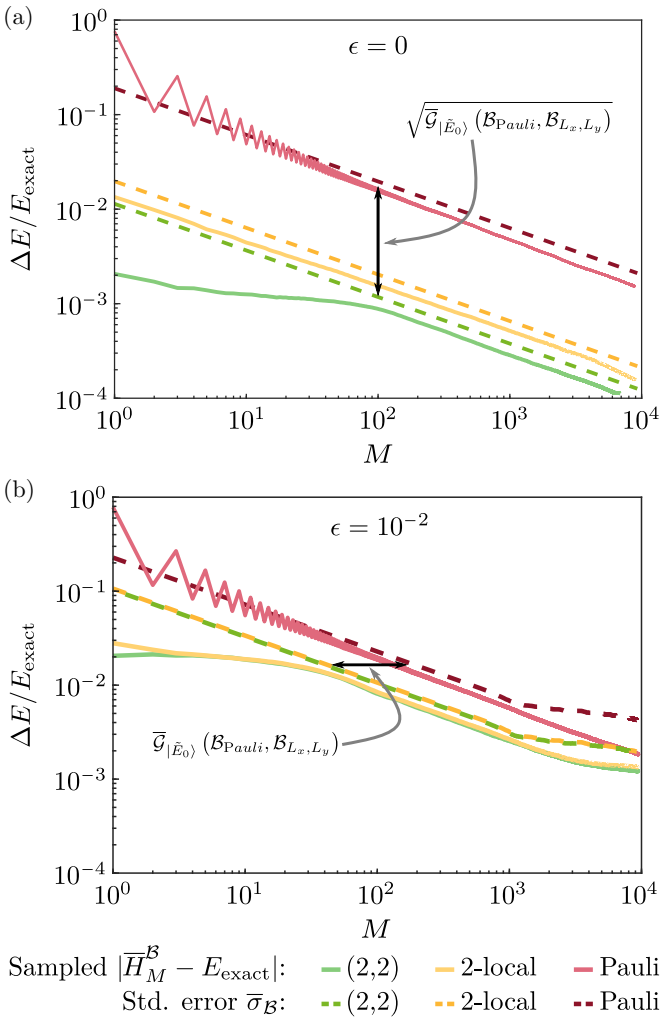

FIG. 8. Numerical example: Sampled energy estimations. The transverse field Ising model with $J/h = 1$ is the same system as the transverse field XY model in Fig. 6 with $\eta = 1$, where we simulate sampling from a perfect $|\psi\rangle = |E_0\rangle$ (a), and a noisy one with global depolarization noise $\epsilon = 1\%$ (b). On real hardware, the noise could have originated from noisy hardware gates, imperfect implementation of the state preparation or of the measurement unitaries U_b. In light red, yellow, and green, we show the relative error of the sampled energy $\overline{H}_M^{\mathcal{B}}$ for Pauli, 2-local, and 2×2 geometric partitioning. To make a fair comparison, we report the mean energy error over a random ordering of the simulated samples. In darker shades of the same colors, we give the corresponding relative standard errors $\overline{\sigma}_{\mathcal{B}}/E_{\text{exact}}$ [compare also Eq. (34)], which match well with the sampled energy estimates. For few samples, Pauli partitioning suffers from its integer spectrum, while the sampled energies with geometric partitioning are more fine-grained and closer to the exact energy. Consistently with Fig. 6(c), we find that even the noisy geometric partitionings in (b) yield lower shot-noise than the noise-free Pauli bases measurements in (a).

B. Simulated sampling

In this section, we give further evidence that the measures for the accuracy of sampling strategies that we use, see Eqs. (8) and (34), are good approximations to the actual error occurring in the sampling process. These results thus support that the starting point for our comparison of sampling efficiency is well-founded. To provide this evidence,

we show simulations of actual sampling. This means we do not compute the considered variances from the full wavefunction but instead simulate actual measurements by applying the measurement unitaries U_{H_b} and sampling bit-strings from which we compute the corresponding energy values. For the analysis of noisy situations, we add random contributions to the state according to Eq. (30) before sampling the bit-strings. For these simulations, we compute the relative energy error from the sampling process

$$\left|\frac{\overline{H}_M^{\mathcal{B}} - E_{\text{exact}}}{E_{\text{exact}}}\right| \tag{56}$$

and compare it to $\sigma_{\mathcal{B}}/|E_{\text{exact}}|$ for the noiseless and $\overline{\sigma}_{\mathcal{B}}/|E_{\text{exact}}|$ for the noisy case.

We do this comparison for the TFXYM and TFIM ground state $|E_0\rangle$ at $J = \eta = h = 1$ on a 4×6 rectangular lattice as examples. Figures 8(a) and 8(b) show the relative energy error under simulated samples in light red for Pauli sampling, light yellow for 2-local sampling, and light green for a $(2,2)$ geometric partition as well as the corresponding values of $\sigma_{\mathcal{B}}/|E_{\text{exact}}|$ for the noiseless case in panel (a) and $\overline{\sigma}_{\mathcal{B}}/|E_{\text{exact}}|$ for the noisy case in panel (b).

A first observation is that the light and darker lines of the same color are indeed very close to each other. As expected, $\sigma_{\mathcal{B}}/|E_{\text{exact}}|$ and $\overline{\sigma}_{\mathcal{B}}/|E_{\text{exact}}|$ are slightly above the corresponding observed mean sampling error (light colors), see Eqs. (8) and (9). For few samples, $M \lessapprox 100$, there are discrete jumps, also above the standard error for Pauli partitioning. This reflects the coarse discretization of the Pauli spectrum, where obtainable one-shot expectation values are integers. On the other hand, for few samples, geometric partitioning performs even better than its standard error. This is consistent with a further observation, namely that only rather few eigenstates of each patch of a geometric partitioning have nonvanishing occupation, leading to a better discrete representation of E_{exact} with very few samples. We conjecture this beneficial few-sample effect also to occur at other examples, making geometric partitioning even more beneficial than quantified by $\overline{\mathcal{G}}_{|\tilde{E}_0\rangle}\left(\mathcal{B}_{\text{Pauli}}, \mathcal{B}_{L_x,L_y}\right)$.

Besides showing the appropriateness of the quantities $\sigma_{\mathcal{B}}$ and $\overline{\sigma}_{\mathcal{B}}$ for quantifying sampling errors, Fig. 8 again confirms the reduction of sampling complexity via geometric partitioning as discussed in Secs. III B and IV A. In particular, Fig. 8(b), where we take errors in the state preparation and measurement circuit implementation into account, confirms that there is no fundamental change in the accuracy of our analysis, while the improvement $\mathcal{G}$ (Definition 1) is reduced to $\overline{\mathcal{G}}$ (Definition 2), as predicted in Theorem 2 and shown in Figs. 6 and 7. In good agreement with Theorem 2, even for a noise level of 1% geometric partitioning, green and yellow in Fig. 8(b), still yield a lower shot-noise than perfect state preparation and noiseless sampling in the Pauli bases, red in Fig. 8(a).

V. DISCUSSION

The main idea we pursued in this work was to make use of the geometric locality of quantum systems to improve measurement strategies compared to sampling in Pauli bases. Conceptually, we think of these geometrically local subsystems to be diagonalizable on the quantum hardware, meaning that the transformations mapping the eigenstates of the subsystems to the computational basis can be found.

In our first main result (Sec. III), we have proven that indeed geometric partitioning always yields an advantage over Pauli partitioning for energy estimates of a system in an energy eigenstate, where we identified three cases. These are geometric partitions coming from a one-dimensional cuts, two-dimensional patches or even 2-local partitions. Typically, smaller guaranteed improvements coincide with easier-to-implement measurement circuits, as these act on less qubits/lattice sites. There is thus something for (almost) any available measurement gate budget, starting from as little as one additional 2-qubit gate per patch when using the 2-local geometric partitioning strategy. Twice as large partitioning patches with, e.g., $(L_x, L_y) = (2,2)$ then require to implement a 4-qubit measurement unitary, which appears still feasible also in near-term applications. Throughout our numerical examples, including different Ising and Hubbard models, we find that typically already 4-qubit $(L_x, L_y) = (2,2)$ partitioning suffices to obtain orders of magnitude relative sampling complexity improvements. In such case, the gate complexity to implement measurement circuit U_{H_b} scales with the number of blocks U_b, and hence only linear in the number of qubits n

With our second main result (Sec. IV), we cover the more general case of imperfect eigenstates, perturbed by global depolarization noise, which can be viewed as a proxy for both imperfect state preparation and quantum hardware noise. A future extension of our work may include tailoring readout circuits not only based on state symmetries as we did here, but also based on a quantum hardware-specific noise model. We identified three noise regimes. In the low noise one (*I*.), the noiseless sampling improvement is maintained up to at worst a factor 2. In the intermediate (*II*.) and high-noise (*III*.) regime, measuring in the geometric basis effectively becomes as good as measuring in the eigenbasis, again up to a factor 2. Importantly, we provide upper bounds on the number of additional gates that can be used for the measurement circuits and the amount of additional noise they may introduce while leading to a sampling advantage for geometric partitioning, see Corollaries 2 and 3.

Typically, more correlated states require larger measurement efforts [48]. Hence, our approach, which measures geometrically extended operators, yields very significant improvements for highly correlated states, which are typically of interest. We observed this behavior featuring improvements $\gtrapprox 10^7$ for TFXYM around its oscillatory

to ferromagnetic phase transition [Fig. 2(a)]. We focused on proving sampling improvements for geometric local models, both in bosonic and fermionic bases. We have shown in total six examples, where the geometric partitioning strategy led to a sampling improvement compared to the (Pauli) baseline, as our analysis suggested. Further, we demonstrated with the transverse field biaxial next-nearest-neighbor Ising model, how this idea extends beyond 2-local models to a quasi-3-local one.

To go forward an interesting future direction might be potentially lifting the requirement of geometric locality, and extending the idea to further types of symmetries beyond translational invariance. Similarly, we found examples of improvements beyond our lower bound, e.g., within the ferromagnetic phase $J/h \gg 1$ of the TFIM. This motivates to investigate further more model-specific sampling improvement estimations or to exploit perturbation theory results to construct better-performing sampling basis as we provided with our Lemma 2.

We took the stand here that we are given a state and aim to estimate the energy expectation value. In future work, one could flip this view around and rather use geometrically local bases to compress Ansätze for approximated state preparation. For instance, for ground state preparation, one could expect the low-energy subspace of the geometrically local bases to be mostly relevant for the ground state of the entire system. Provided the inverses of our measurement unitaries not only lead to vanishing local energy variances but also low local energy expectation values, this strategy would support approximations with reduced effective Hilbert spaces.

ACKNOWLEDGMENTS

This work was supported by German Research Foundation (DFG)—Project-ID 429529648—TRR 306 QuCoLiMa, the German Federal Ministry of Education and Research (BMBF) Contract No. 13N16067 "EQUAHUMO," and Munich Quantum Valley, which is supported by the Bavarian state government with funds from the Hightech Agenda Bayern Plus. This work used resources of the Erlangen National High Performance Computing Center. T.E. acknowledges support from the International Max-Planck Research School for Physics of Light. M.K. is co-funded by the ERDF of the European Union and by Fonds of the Hamburg Ministry of Science, Research, Equalities and Districts (BWFGB) and funded by the Fujitsu Germany GmbH as part of the endowed professorship "Quantum Inspired and Quantum Optimization."

DATA AVAILABILITY

The data that support the findings of this article are not publicly available. The data are available from the authors upon reasonable request.

APPENDIX A: HOW TO COMPARE MEASUREMENT STRATEGIES [PROOF OF EQ. (8)]

A measurement strategy in quantum mechanics has the aim of determining an estimator for the true observable expectation value $\langle\psi|H|\psi\rangle$ as accurate as possible with a given number of measurements M. Then standard deviation of an unbiased estimator is given by

$$\sigma^2_{\text{min.}} = \frac{1}{M}\Delta H^2 = \frac{1}{M}\text{Var}_{|\psi\rangle}(H). \tag{A1}$$

Realizing this sampling error requires the ability to measure observable H at once. However, finding the eigenbasis, in which one would need to measure, generally cannot be done in polynomial time, as we need to diagonalize the entire exponentially large Hilbert space to map $|\psi\rangle$ into the eigenbasis of H, in the worst case. Thus, in typical nontrivial sampling tasks, one needs to split H into parts H_b of which the transformations into their respective eigenbasis are known, such that $H = \sum_{H_b\in\mathcal{B}} H_b$ with partitioning $\mathcal{B}$. Then the measurement error can be upper bounded by [step by step calculation of Eq. (10)],

$$\frac{1}{M}\Delta H^2 = \frac{1}{M}\left[\sum_{H_b\in\mathcal{B}}\text{Var}_{|\psi\rangle}(H_b) + \sum_{\substack{H_b,H_b'\in\mathcal{B}\\ H_b\neq H_b'}}\text{CoV}_{|\psi\rangle}(H_b,H_b')\right] \leq \frac{1}{M}\left[\sum_{H_b\in\mathcal{B}}\text{Var}_{|\psi\rangle}(H_b) + \sum_{\substack{H_b,H_b'\in\mathcal{B}\\ H_b\neq H_b'}}\left|\text{CoV}_{|\psi\rangle}(H_b,H_b')\right|\right]$$
$$\leq \frac{1}{M}\left[\sum_{H_b\in\mathcal{B}}\text{Var}_{|\psi\rangle}(H_b) + \sum_{\substack{H_b,H_b'\in\mathcal{B}\\ H_b\neq H_b'}}\sqrt{\text{Var}_{|\psi\rangle}(H_b)\,\text{Var}_{|\psi\rangle}(H_b')}\right] = \frac{1}{M}\left[\sum_{H_b\in\mathcal{B}}\sqrt{\text{Var}_{|\psi\rangle}(H_b)}\right]^2 \leq \sum_{H_b\in\mathcal{B}}\frac{\text{Var}_{|\psi\rangle}(H_b)}{M_b}, \tag{A2}$$

where $M = \sum_{b=1}^{|\mathcal{B}|} M_b$. The first inequality adds absolute values to the covariances and the second inequality is the Cauchy-Schwarz inequality. Using Lagrangian multipliers, one finds that the last inequality becomes an equality precisely for the optimal distribution of sampling budget M. Let us write this down explicitly. The Lagrangian is given by

$$\mathcal{L}(\vec{M}, \lambda) := \sum_{b=1}^{|\mathcal{B}|} \frac{\mathrm{Var}_{|\psi\rangle}(H_b)}{M_b} - \lambda \left(M - \sum_{b=1}^{|\mathcal{B}|} M_b \right), \quad \text{(A3)}$$

where we named $\vec{M}$ being a vector of all M_b. Then the stationary conditions yields

$$0 \stackrel{!}{=} \frac{\mathcal{L}(\vec{M}, \lambda)}{\partial M_b} = -\frac{\mathrm{Var}_{|\psi\rangle}(H_b)}{M_b^2} + \lambda \quad \forall b \in \{1, 2, \ldots, B\}, \quad \text{(A4)}$$

$$0 \stackrel{!}{=} \frac{\mathcal{L}(\vec{M}, \lambda)}{\partial \lambda} = M - \sum_{b=1}^{B} M_b, \quad \text{(A5)}$$

where $\stackrel{!}{=}$ means that we demand this equality to be fulfilled for the stationary conditions. As $\mathrm{Var}_{|\psi\rangle}(H_b) \geq 0$, we know that $\lambda = \mathrm{Var}_{|\psi\rangle}(H_b)/M_b^2 \geq 0$, and in order to minimize the standard error, M_b needs to be chosen such that all $\sqrt{\mathrm{Var}_{|\psi\rangle}(H_b)}/M_b$ become equal. This allows us to express λ in terms of M and $\mathrm{Var}_{|\psi\rangle}(H_b)$,

$$M = \sum_{b=1}^{|\mathcal{B}|} M_b = \sum_{b=1}^{|\mathcal{B}|} \sqrt{\frac{\mathrm{Var}_{|\psi\rangle}(H_b)}{\lambda}}$$

$$\Rightarrow \quad \sqrt{\lambda} = \frac{1}{M} \sum_{b=1}^{|\mathcal{B}|} \sqrt{\mathrm{Var}_{|\psi\rangle}(H_b)}. \quad \text{(A6)}$$

Then Eqs. (A6) and (A4) solve the constraint minimization problem,

$$\min_{M=\sum M_b} \left(\sum_{H_b \in \mathcal{B}} \frac{\mathrm{Var}_{|\psi\rangle}(H_b)}{M_b} \right) = \sum_{H_b \in \mathcal{B}} \sqrt{\mathrm{Var}_{|\psi\rangle}(H_b)} \sqrt{\lambda} = \frac{1}{M} \left[\sum_{H_b \in \mathcal{B}} \sqrt{\mathrm{Var}_{|\psi\rangle}(H_b)} \right]^2. \quad \text{(A7)}$$

$\mathrm{Var}_{|\psi\rangle}(H_b)$ are the sampling errors we will experience when measuring H in the partitioning $\mathcal{B}$. The partitioning $\mathcal{B}$ is by no means unique. Rather it gives us a way to compare sampling budget allocation agnostic sampling strategies by comparing $\left[\sum_{H_b \in \mathcal{B}} \sqrt{\mathrm{Var}_{|\psi\rangle}(H_b)}\right]^2$, motivation our relative sampling complexity cost function in Definition 1.

In a more general setting, where we are interested to upper bound the (average) measurement error, not only with respect to a single state $|\psi\rangle$, but with respect to an assemble of states $|\tilde{\psi}\rangle$, we need to consider the expectation value over $|\tilde{\psi}\rangle$,

$$\frac{1}{M} \overline{\Delta H}^2 = \frac{1}{M} \mathbb{E} \left[\mathrm{Var}_{|\tilde{\psi}\rangle}(H) \right]. \quad \text{(A8)}$$

Applying once more the Cauchy-Schwarz inequality, this time on the expectation value over $|\tilde{\psi}\rangle$, as well as using linearity of the expectation value, we can obtain a similar upper bound to Eq. (A2),

$$\begin{aligned}
\frac{1}{M} \overline{\Delta H}^2 = \frac{1}{M} \mathbb{E} \left[\mathrm{Var}_{|\tilde{\psi}\rangle}(H) \right] &\leq \frac{1}{M} \mathbb{E} \left[\left[\sum_{H_b \in \mathcal{B}} \sqrt{\mathrm{Var}_{|\psi\rangle}(H_b)} \right]^2 \right] \\
&= \frac{1}{M} \mathbb{E} \left[\sum_{H_b \in \mathcal{B}} \mathrm{Var}_{|\psi\rangle}(H_b) \right] + \frac{1}{M} \mathbb{E} \left[\sum_{H_b' \neq H_b \in \mathcal{B}} \sqrt{\mathrm{Var}_{|\psi\rangle}(H_b)} \sqrt{\mathrm{Var}_{|\psi\rangle}(H_b')} \right] \\
&\leq \frac{1}{M} \sum_{H_b \in \mathcal{B}} \mathbb{E} \left[\mathrm{Var}_{|\psi\rangle}(H_b) \right] + \frac{1}{M} \sum_{H_b' \neq H_b \in \mathcal{B}} \sqrt{\mathbb{E} \left[\mathrm{Var}_{|\psi\rangle}(H_b) \right]} \sqrt{\mathbb{E} \left[\mathrm{Var}_{|\psi\rangle}(H_b') \right]} \\
&= \frac{1}{M} \left[\sum_{H_b \in \mathcal{B}} \sqrt{\mathbb{E} \left[\mathrm{Var}_{|\tilde{\psi}\rangle}(H_b) \right]} \right]^2 = \sigma_{\mathcal{B}}^2. \quad \text{(A9)}
\end{aligned}$$

To obtain the state-independent average sampling error, one can average Eq. (A9) over all states $|\tilde{\psi}\rangle$. Lemma 3 covers this case for $\epsilon = 1$, $|\tilde{\psi}\rangle = |\xi\rangle$, and $|\xi\rangle$ Haar-random up to second moment,

$$\frac{1}{M}\overline{\Delta H}^2 \leq \frac{1}{M}\left[\sum_{H_b\in\mathcal{B}}\sqrt{\mathbb{E}\left[\mathrm{Var}_{|\xi\rangle}\left(H_b\right)\right]}\right]^2 = \frac{1}{M}\left[\sum_{H_b\in\mathcal{B}}\sqrt{\|H_b\|_F^2/d}\right]^2 + \mathcal{O}\left(\frac{1}{d}\right) := \sigma_{\text{av.}}^2 . \tag{A10}$$

Note that shadow norms are closely related to the Frobenius norm, also called Hilbert-Schmidt norm [10]. Similarly, one can derive a worst-case sampling error from Eq. (A2), by partitioning $H = \sum_b \omega_b P_b$ into Pauli weights ω_b and Pauli strings P_b. Then the each variance $\mathrm{Var}(P_b) = \langle P_b^2\rangle - \langle P_b\rangle^2 = 1 - \langle P_b\rangle^2$ is upper bound by 1, and we get

$$\frac{1}{M}\Delta H^2 \leq \frac{1}{M}\left[\sum_{H_b\in\mathcal{B}}\sqrt{\mathrm{Var}_{|\psi\rangle}\left(H_b\right)}\right]^2 = \frac{1}{M}\left[\sum_{H=\sum_b\omega_b P_b}\sqrt{\omega_b^2\,\mathrm{Var}_{|\psi\rangle}\left(P_b\right)}\right]^2 \leq \frac{1}{M}\left[\sum_{H=\sum_b\omega_b P_b}|\omega_b|\right]^2 := \sigma_{\text{max.}}^2 . \tag{A11}$$

APPENDIX B: NOISELESS SAMPLING IMPROVEMENTS: ADDITIONAL MATERIAL FOR SEC. III

1. Definitions of employed partitionings

Here, we write out explicitly all used partitionings. To recall Eq. (5), the considered Hamiltonians have following form:

$$H = \sum_{i,\alpha} h_i^{(\alpha)} o_i^{(\alpha)} + \sum_{\langle i,j\rangle,\alpha,\beta} J_{i,j}^{(\alpha,\beta)} o_i^{(\alpha)} o_j^{(\beta)}, \tag{B1}$$

with $o_i^{(\alpha)}$ Pauli or annihilation/creation operators.

Definition B1 (Mutually commuting operator/Pauli partitioning). Let H be a lattice Hamiltonian as defined in Eq. (5), which we denote for simplicity by

$$H = \sum_{\iota=1}^{I} O_\iota, \tag{B2}$$

with finite multi-index ι mapped to the natural numbers, and $O_\iota \in \{h_i^{(\alpha)} o_i^{(\alpha)}, J_{i,j}^{(\alpha,\beta)} o_i^{(\alpha)} o_j^{(\beta)}\}$. As H is a 2-local Hamiltonian, $I = \mathcal{O}(n)$. We define the mutually commuting operator partitioning by the following algorithm:

ALGORITHM 1. For construction of a mutually commuting operator partitioning.

$H_1 \leftarrow O_1$;
for $\iota \leftarrow 2$ **to** I **do**
  $b \leftarrow 1$;
  while $\exists O' \in H_b : [O', O_\iota] \neq 0$ **do**
    $b \leftarrow b + 1$;
  end
  $H_b \leftarrow H_b + O_\iota$;
end

where $b \leftarrow b+1$ means that variable b is assigned value $b+1$. We call the resulting partitioning $\mathcal{B}_{\text{Pauli}} = \{H_b\}$ interchangeably Pauli partitioning, as it coincides for most local lattice models with naively separating H by local Pauli operator type.

This way, each H_b can be measured at once. The maximal number of Pauli partitions $|\mathcal{B}_{\text{Pauli}}|$ is given by the number of mutually unbiased bases of a k-local observable $2^k + 1$, with locality $k = 2$, $|\mathcal{B}_{\text{Pauli}}| \leq 5$ [49].

Similarly, we define geometric partitioning explicitly as follows:

Definition B2 (Explicit geometric partitionings). Let H be a lattice Hamiltonian as defined in Eq. (5). We define its geometric partitioning, $\mathcal{B}_{L_x,L_y}$, $H_1 + H_2 = H$ by

$$H_1 := \frac{1}{2}\left(H - H_{\text{cut}} + H'_{\text{cut}}\right), \tag{B3}$$

$$H_2 := \frac{1}{2}\left(H + H_{\text{cut}} - H'_{\text{cut}}\right), \tag{B4}$$

with H_{cut} being the 2-local terms of H cut in H_1 and added to H_2, and vice versa H'_{cut} to obtain subsystem patches $H_{b,k}$ with disjoint support. We use the notation

$$\mathcal{H}_i = \sum_\alpha h_i^{(\alpha)} o_i^{(\alpha)} \quad \text{and} \quad \mathcal{H}_{i,j} = \sum_{\alpha,\beta} J_{i,j}^{(\alpha,\beta)} o_i^{(\alpha)} o_j^{(\beta)}, \tag{B5}$$

for the one-local and two-local terms in Eq. (5), where the site indices i and j have two components for two-dimensional lattices, $i = (i_x, i_y)$ and $j = (j_x, j_y)$, and consider the following three specific partitionings.

a. 1D partitioning. with 1D strip thickness L:

$$L=1: \quad H_1=\sum_{m=1}^{\lfloor n_x/2\rfloor}\sum_{\ell=1}^{n_y}\mathcal{H}_{(2m,\ell),(2m+1,\ell)}+\frac{1}{2}\sum_i\mathcal{H}_i, \quad H_2=\sum_{m=1}^{n_x}\sum_{\ell=1}^{\lfloor n_y/2\rfloor}\mathcal{H}_{(m,2\ell),(m,2\ell+1)}+\frac{1}{2}\sum_i\mathcal{H}_i, \tag{B6}$$

$$L\geq 2: \quad H_{\text{cut}}=\sum_{m=1}^{\lfloor n_x/L\rfloor}\sum_{\ell=1}^{n_y}\mathcal{H}_{(mL,\ell),(mL+1,\ell)}, \qquad H'_{\text{cut}}=\sum_{m=1}^{\lfloor n_x/L\rfloor}\sum_{\ell=1}^{n_y}\mathcal{H}_{(mL-1,\ell),(mL,\ell)}, \tag{B7}$$

where $L=1$ is a special case with H_{cut} consisting of all 2-local terms in x-direction and respectively H'_{cut} containing all 2-local terms in y-direction.

b. 2D partitioning. with patch size $L_x\times L_y$:

$$H_{\text{cut}}=H^x_{\text{cut}}+H^y_{\text{cut}}, \quad H^x_{\text{cut}}=\sum_{m=1}^{\lfloor n_x/L_x\rfloor}\sum_{\ell=1}^{n_y}\mathcal{H}_{(mL_x,\ell),(mL_x+1,\ell)}, \quad H^y_{\text{cut}}=\sum_{m=1}^{n_x}\sum_{\ell=1}^{\lfloor n_y/L_y\rfloor}\mathcal{H}_{(m,\ell L_y),(m,\ell L_y+1)}, \tag{B8}$$

$$H'_{\text{cut}}=H^{x'}_{\text{cut}}+H^{y'}_{\text{cut}}, \quad H^{x'}_{\text{cut}}=\sum_{m=1}^{\lfloor n_x/L_x\rfloor}\sum_{\ell=1}^{n_y}\mathcal{H}_{(mL_x-1,\ell),(mL_x,\ell)}, \quad H^{y'}_{\text{cut}}=\sum_{m=1}^{n_x}\sum_{\ell=1}^{\lfloor n_y/L_y\rfloor}\mathcal{H}_{(m,\ell L_y-1),(m,\ell L_y)}. \tag{B9}$$

c. 2-local partitioning. Further, we can partition H into H_b's with disjoint blocks of size 2, which requires four parts $H=H_1+H_2+H_3+H_4$,

$$H_1=\sum_{m=1}^{\lfloor n_x/2\rfloor}\sum_{\ell=1}^{n_y}\mathcal{H}_{(2m,\ell),(2m+1,\ell)}+\frac{1}{4}\sum_i\mathcal{H}_i, \tag{B10}$$

$$H_2=\sum_{m=1}^{\lfloor n_x/2\rfloor}\sum_{\ell=1}^{n_y}\mathcal{H}_{(2m-1,\ell),(2m,\ell)}+\frac{1}{4}\sum_i\mathcal{H}_i, \tag{B11}$$

$$H_3=\sum_{m=1}^{n_x}\sum_{\ell=1}^{\lfloor n_y/2\rfloor}\mathcal{H}_{(m,2\ell),(m,2\ell+1)}+\frac{1}{4}\sum_i\mathcal{H}_i, \tag{B12}$$

$$H_4=\sum_{m=1}^{n_x}\sum_{\ell=1}^{\lfloor n_y/2\rfloor}\mathcal{H}_{(m,2\ell-1),(m,2\ell)}+\frac{1}{4}\sum_i\mathcal{H}_i. \tag{B13}$$

On the following pages, we provide explicitly all partitions of the six numerical examples of the main text.

2. Relative sampling complexity improvement lower bound (Proof of Theorem 1)

To prove a sampling improvement for measuring in the eigenbases of decoupled subsystems, we express the decoupled Hamiltonians as an orbit of a subgroup of the translations along the x- and y-directions of a two-dimensional lattice. The operators T^x_ℓ and T^y_ℓ denote a translation by ℓ sites into the respective direction. Both directions are connected via a local rotation operation S that maps $T^y_\ell=ST^x_\ell S$. Altogether, we consider subgroups of the symmetry group

$$\mathcal{T}=\{T^x_\ell T^y_m|\ell=1,\dots,n_x,m=1,\dots,n_y\}. \tag{B14}$$

With this, every k-local, translation-invariant Hamiltonian can be written as an orbit of $\mathcal{T}$ acting on a unit cell interaction V, i.e.,

$$H=\sum_{T\in\mathcal{T}}TVT^\dagger+H_{\text{1-local}}. \tag{B15}$$

Here, we focus on geometrically local lattice Hamiltonians as given in Eq. (5) so that $H_{\text{1-local}}=\sum_{i,\alpha}h_i^{(\alpha)}o_i^{(\alpha)}$ and $V=\sum_{\alpha,\beta}J^{(\alpha,\beta)}_{i_0,j_0}o^{(\alpha)}_{i_0}o^{(\beta)}_{j_0}$ with arbitrary neighboring lattice vertices i_0 and j_0. We will use in the following that any Pauli partitioning $\mathcal{B}_{\text{Pauli}}$ consists of at least two partitions. Further, at least one of them needs to have translationally invariant 2-local terms. Therefore, assuming all 2-local terms by themselves and respectively all 1-local terms by themselves can be measured simultaneously, and yields a lower bound on the Pauli sampling error,

$$\begin{aligned}\sigma^2_{\text{Pauli}}&=\frac{1}{M}\left[\sum_{H_b\in\mathcal{B}_{\text{Pauli}}}\sqrt{\text{Var}_{|E_i\rangle}(H_b)}\right]^2\\&\geq\frac{2}{M}\left[\text{Var}_{|E_i\rangle}(H_{\text{2-local}})+\text{Var}_{|E_i\rangle}(H_{\text{1-local}})\right].\end{aligned} \tag{B16}$$

Further, as we focus on eigenstates here, we have $\text{Var}_{|E_i\rangle}(H_{\text{2-local}})=\text{Var}_{|E_i\rangle}(H_{\text{1-local}})$ due to Corollary 1. We

can write the horizontal interactions as an orbit of the translation group acting on a unit cell Hamiltonian V,

$$H^h_{2\text{-local}} = \sum_{\ell=1}^{n_x}\sum_{m=1}^{n_y} T^y_m T^x_\ell V T^x_{-\ell} T^y_{-m}, \tag{B17}$$

and the vertical interactions accordingly

$$H^v_{2\text{-local}} = S H^h_{2\text{-local}} S. \tag{B18}$$

We begin with a proof for 2D partitioning (compare Definition B2 b.). The sample improvement reads

$$\mathcal{G}_{|E_i\rangle}\left(\mathcal{B}_{\text{Pauli}}, \mathcal{B}_{L_x,L_y}\right) \geq \frac{\text{Var}_{|E_i\rangle}(H_{2\text{-local}})}{\text{Var}_{|E_i\rangle}\left(\frac{1}{2}(H - H^x_{\text{cut}} - H^y_{\text{cut}} + H^{x'}_{\text{cut}} + H^{y'}_{\text{cut}})\right)}. \tag{B19}$$

Let us consider the numerator and denominator separately. The denominator can be simplified further via

$$\begin{aligned}&\text{Var}_{|E_i\rangle}\left(\frac{1}{2}(H - H^x_{\text{cut}} - H^y_{\text{cut}} + H^{x'}_{\text{cut}} + H^{y'}_{\text{cut}})\right)\\ &= \text{Var}_{|E_i\rangle}\left(\frac{1}{2}(H + H^x_{\text{cut}} + H^y_{\text{cut}} - H^{x'}_{\text{cut}} - H^{y'}_{\text{cut}})\right)\\ &= \frac{1}{4}\,\text{Var}_{|E_i\rangle}\left(H^x_{\text{cut}} + H^y_{\text{cut}} - H^{x'}_{\text{cut}} - H^{y'}_{\text{cut}}\right).\end{aligned} \tag{B20}$$

First, observe that $\text{Var}_{|E_i\rangle}(H_{2\text{-local}}) = 2\,\text{Var}_{|E_i\rangle}(H^h_{2\text{-local}}) + 2\,\text{Var}_{|E_i\rangle}(SH^h_{2\text{-local}})$ from invariance of the ground state. Further, we can express both terms as covariances. We have

$$\begin{aligned}&\text{Var}_{|E_i\rangle}(H^h_{2\text{-local}})\\ &= \text{Var}_{|E_i\rangle}\left(\sum_{\ell=1}^{n_x}\sum_{m=1}^{n_y} T^y_m T^x_\ell V T^x_{-\ell} T^y_{-m}\right)\\ &= n_x n_y \sum_{\ell=1}^{n_x}\sum_{m=1}^{n_y} \text{Var}_{|E_i\rangle}(T^y_m T^x_\ell V),\end{aligned} \tag{B21}$$

where we used the multilinearity of the covariance, the fact that translation operators commute, $\left[T^x_\ell, T^y_m\right] = 0$, and again symmetry of the ground state to identify $n_x n_y$ identical terms. Similarly, we can calculate

$$\text{Var}_{|E_i\rangle}(SH^h_{2\text{-local}}) = n_x n_y \sum_{\ell=1}^{n_x}\sum_{m=1}^{n_y} \text{Var}_{|E_i\rangle}(ST^y_m T^x_\ell V). \tag{B22}$$

Let us separate the qubit number $n = n_x n_y$ from the correlators and write $\text{Var}_{|E_i\rangle}(H_{2\text{-local}}) =: 2nC_1$. The denominator can be treated similarly

$$\begin{aligned}&\text{Var}_{|E_i\rangle}(H^x_{\text{cut}} + H^y_{\text{cut}} - H^{x'}_{\text{cut}} - H^{y'}_{\text{cut}})\\ &= 2\,\text{Var}_{|E_i\rangle}(H^x_{\text{cut}}) + 2\,\text{Var}_{|E_i\rangle}(H^y_{\text{cut}}) - \text{Var}_{|E_i\rangle}(T^x_1 H^x_{\text{cut}})\\ &\quad - \text{Var}_{|E_i\rangle}(T^x_{-1} H^x_{\text{cut}}) - \text{Var}_{|E_i\rangle}(T^y_1 H^y_{\text{cut}})\\ &\quad - \text{Var}_{|E_i\rangle}(T^y_{-1} H^y_{\text{cut}}).\end{aligned} \tag{B23}$$

Expressing H^x_{cut} in terms of V, we calculate

$$\begin{aligned}\text{Var}_{|E_i\rangle}(H^x_{\text{cut}}) &= \text{Var}_{|E_i\rangle}\left(\sum_{\ell=1}^{n_x/L_x}\sum_{m=1}^{n_y} T^x_{\ell L_x} T^y_m V T^y_{-m} T^x_{-\ell L_x}\right)\\ &= n_y \frac{n_x}{L_x}\sum_{\ell=1}^{n_x/L_x}\sum_{m=1}^{n_y} \text{Var}_{|E_i\rangle}\left(T^x_{\ell L_x} T^y_m V\right),\end{aligned} \tag{B24}$$

and similarly,

$$\text{Var}_{|E_i\rangle}(H^y_{\text{cut}}) = n_x \frac{n_y}{L_y}\sum_{m=1}^{n_y/L_y}\sum_{\ell=1}^{n_x} \text{Var}_{|E_i\rangle}\left(T^y_{mL_y} T^x_\ell S V\right), \tag{B25}$$

$$\text{Var}_{|E_i\rangle}(T^x_1 H^x_{\text{cut}}) = n_y \frac{n_x}{L_x}\sum_{\ell=1}^{n_x/L_x}\sum_{m=1}^{n_y} \text{Var}_{|E_i\rangle}\left(T^x_{\ell L_x+1} T^y_m V\right), \tag{B26}$$

$$\text{Var}_{|E_i\rangle}(T^x_{-1} H^x_{\text{cut}}) = n_y \frac{n_x}{L_x}\sum_{\ell=1}^{n_x/L_x}\sum_{m=1}^{n_y} \text{Var}_{|E_i\rangle}\left(T^x_{\ell L_x-1} T^y_m V\right), \tag{B27}$$

$$\text{Var}_{|E_i\rangle}(T^y_1 H^y_{\text{cut}}) = n_x \frac{n_y}{L_y}\sum_{m=1}^{n_y/L_y}\sum_{\ell=1}^{n_x} \text{Var}_{|E_i\rangle}\left(T^y_{mL_y+1} T^x_\ell S V\right). \tag{B28}$$

Separating qubit number and cluster thickness, we write $\text{Var}_{|E_i\rangle}\left(H^x_{\text{cut}} + H^y_{\text{cut}} - H^{x'}_{\text{cut}} - H^{y'}_{\text{cut}}\right) = 2\left(\frac{n}{L_x}C^x_2 + \frac{n}{L_y}C^y_2\right)$ and altogether

$$\mathcal{G}_{|E_i\rangle}\left(\mathcal{B}_{\text{Pauli}}, \mathcal{B}_{L_x,L_y}\right) = 4L_x L_y \frac{C_1}{L_y C^x_2 + L_x C^y_2}, \tag{B29}$$

with the system-dependent constants

$$C_1 = \sum_{\ell=1}^{n_x}\sum_{m=1}^{n_y} \text{Var}_{|E_i\rangle}(T^y_m T^x_\ell V) + \sum_{\ell=1}^{n_x}\sum_{m=1}^{n_y} \text{Var}_{|E_i\rangle}(ST^y_m T^x_\ell V), \tag{B30}$$

$$C_2^x = \sum_{\ell=1}^{n_x/L_x}\sum_{m=1}^{n_y}\left(\mathrm{Var}_{|E_i\rangle}\left(T^x_{\ell L_x}T^y_m V\right) - \frac{1}{2}\mathrm{Var}_{|E_i\rangle}\left(T^x_{\ell L_x+1}T^y_m V\right) - \frac{1}{2}\mathrm{Var}_{|E_i\rangle}\left(T^x_{\ell L_x-1}T^y_m V\right)\right), \tag{B31}$$

$$C_2^y = \sum_{m=1}^{n_y/L_y}\sum_{\ell=1}^{n_x}\left(\mathrm{Var}_{|E_i\rangle}\left(T^y_{mL_y}T^x_\ell SV\right) - \frac{1}{2}\mathrm{Var}_{|E_i\rangle}\left(T^y_{mL_y+1}T^x_\ell SV\right) - \frac{1}{2}\mathrm{Var}_{|E_i\rangle}\left(T^y_{mL_y-1}T^x_\ell SV\right)\right). \tag{B32}$$

From $C_2^x \le C_1$ and $C_2^y \le C_1$, we already have $\mathcal{G}_{|E_i\rangle}\left(\mathcal{B}_{\mathrm{Pauli}}, \mathcal{B}_{L_x,L_y}\right) \ge (4L_xL_y)/(L_x+L_y)$ as a lower bound which can be further improved. We have, similarly to before:

$$H_{\mathrm{cut}} = T^y_{-1}T^x_{-1}H'_{\mathrm{cut}}T^y_1T^x_1 \quad \Rightarrow \quad \mathrm{Var}_{|E_i\rangle}\left(H_{\mathrm{cut}}\right) = \mathrm{Var}_{|E_i\rangle}\left(H'_{\mathrm{cut}}\right) \tag{B33}$$

due to the translational invariance of the eigenstate. Thus, we can write the denominator as

$$\mathrm{Var}_{|E_i\rangle}\left(H^x_{\mathrm{cut}} + H^y_{\mathrm{cut}} - H^{x'}_{\mathrm{cut}} - H^{y'}_{\mathrm{cut}}\right) = \mathrm{Var}_{|E_i\rangle}\left(H_{\mathrm{cut}} - H'_{\mathrm{cut}}\right) \tag{B34}$$

$$= \mathrm{Var}_{|E_i\rangle}\left(H_{\mathrm{cut}}\right) + \mathrm{Var}_{|E_i\rangle}\left(H'_{\mathrm{cut}}\right) - 2\mathrm{CoV}_{|E_i\rangle}\left(H_{\mathrm{cut}}, H'_{\mathrm{cut}}\right) \tag{B35}$$

$$= 2\,\mathrm{Var}_{|E_i\rangle}\left(H_{\mathrm{cut}}\right)\left[1 - \frac{\mathrm{CoV}_{|E_i\rangle}\left(H_{\mathrm{cut}}, H'_{\mathrm{cut}}\right)}{\sqrt{\mathrm{Var}_{|E_i\rangle}\left(H_{\mathrm{cut}}\right)}\sqrt{\mathrm{Var}_{|E_i\rangle}\left(H'_{\mathrm{cut}}\right)}}\right] \tag{B36}$$

$$= 2\,\mathrm{Var}_{|E_i\rangle}\left(H_{\mathrm{cut}}\right)\left[1 - \mathrm{CoR}_{|E_i\rangle}\left(H_{\mathrm{cut}}, H'_{\mathrm{cut}}\right)\right], \tag{B37}$$

which we simplified by using the correlation $\mathrm{CoR}_{|E_i\rangle}(H_{\mathrm{cut}}, H'_{\mathrm{cut}}) = \mathrm{CoV}_{|E_i\rangle}(H_{\mathrm{cut}}, H'_{\mathrm{cut}})/\sqrt{\mathrm{Var}_{|E_i\rangle}(H_{\mathrm{cut}})\,\mathrm{Var}_{|E_i\rangle}(H'_{\mathrm{cut}})}$. Again due to the translational symmetry of $|E_i\rangle$ we have

$$\mathrm{CoV}_{|E_i\rangle}\left(H_{\mathrm{cut}}, H'_{\mathrm{cut}}\right) = \mathrm{CoV}_{|E_i\rangle}\left(H_{\mathrm{cut}}, T^y_{-1}T^x_{-1}H_{\mathrm{cut}}T^y_1T^x_1\right) = \mathrm{CoV}_{|E_i\rangle}\left(T^x_{-1}T^y_{-1}H_{\mathrm{cut}}, T^y_{-1}T^x_{-1}H_{\mathrm{cut}}\right) \tag{B38}$$

$$= \mathrm{Var}_{|E_i\rangle}\left(T^x_{-1}T^y_{-1}H_{\mathrm{cut}}\right) \ge 0. \tag{B39}$$

From a similar calculation we have $\mathrm{Var}_{|E_i\rangle}\left(H_{\mathrm{cut}}\right) = \mathrm{Var}_{|E_i\rangle}\left(H^x_{\mathrm{cut}} + H^y_{\mathrm{cut}}\right) = \mathrm{Var}_{|E_i\rangle}\left(H^x_{\mathrm{cut}}\right) + \mathrm{Var}_{|E_i\rangle}\left(H^y_{\mathrm{cut}}\right) + 2\mathrm{CoV}_{|E_i\rangle}\left(H^x_{\mathrm{cut}}, H^y_{\mathrm{cut}}\right) \le \mathrm{Var}_{|E_i\rangle}\left(H^x_{\mathrm{cut}}\right) + \mathrm{Var}_{|E_i\rangle}\left(H^y_{\mathrm{cut}}\right)$,

$$\mathcal{G}_{|E_i\rangle}\left(\mathcal{B}_{\mathrm{Pauli}}, \mathcal{B}_{L_x,L_y}\right) \ge 4\frac{\mathrm{Var}_{|E_i\rangle}(H_{2\text{-local}})}{2\,\mathrm{Var}_{|E_i\rangle}\left(H_{\mathrm{cut}}\right)}\frac{1}{1 - \mathrm{CoR}_{|E_i\rangle}\left(H_{\mathrm{cut}}, H'_{\mathrm{cut}}\right)} \tag{B40}$$

$$\ge 4\frac{2nC_1}{2\,\mathrm{Var}_{|E_i\rangle}\left(H^x_{\mathrm{cut}}\right) + 2\,\mathrm{Var}_{|E_i\rangle}\left(H^y_{\mathrm{cut}}\right)}\frac{1}{1 - \mathrm{CoR}_{|E_i\rangle}\left(H_{\mathrm{cut}}, H'_{\mathrm{cut}}\right)} \tag{B41}$$

$$= 4\frac{2nC_1}{2\left(\frac{n}{L_x}C_2^x + \frac{n}{L_y}C_2^y\right)}\frac{1}{1 - \mathrm{CoR}_{|E_i\rangle}\left(H_{\mathrm{cut}}, H'_{\mathrm{cut}}\right)} \tag{B42}$$

$$\ge 4\frac{L_xL_y}{L_x+L_y}\frac{1}{1 - \mathrm{CoR}_{|E_i\rangle}\left(H_{\mathrm{cut}}, H'_{\mathrm{cut}}\right)}. \tag{B43}$$

For 1D partitioning (compare Definition B2 a), we set $L_x = n_x$ and call $L_y =: L$. The sample improvement reads

$$\mathcal{G}_{|E_i\rangle}\left(\mathcal{B}_{\mathrm{Pauli}}, \mathcal{B}_{L_x,L_y}\right) \ge 4\frac{\mathrm{Var}_{|E_i\rangle}(H_{2\text{-local}})}{\mathrm{Var}_{|E_i\rangle}\left(H_{\mathrm{cut}} - H'_{\mathrm{cut}}\right)}. \tag{B44}$$

While the numerator is the same as before, the denominator becomes $\mathrm{Var}_{|E_i\rangle}\left(H_{\mathrm{cut}} - H'_{\mathrm{cut}}\right) = 2\left(\mathrm{Var}_{|E_i\rangle}\left(H_{\mathrm{cut}}\right) - \mathrm{Var}_{|E_i\rangle}\left(T^y_1H_{\mathrm{cut}}\right)\right)$, cf. Eq. (B23). We can further simplify Eq. (B21) for 1D cuts to

$$\mathrm{Var}_{|E_i\rangle}(H^h_{2\text{-local}}) = L\sum_{m=1}^{L}\mathrm{Var}_{|E_i\rangle}(T^y_m H_{\mathrm{cut}}). \tag{B45}$$

Altogether we have

$$\mathcal{G}_{|E_i\rangle}\left(\mathcal{B}_{\text{Pauli}},\mathcal{B}_{L_x,L_y}\right) \geq 4L\frac{\sum_{m=1}^{L}\text{Var}_{|E_i\rangle}(T_m^y H_{\text{cut}})}{\text{Var}_{|E_i\rangle}\left(H_{\text{cut}}\right)-\text{Var}_{|E_i\rangle}\left(T_1^y H_{\text{cut}}\right)} \geq 4L\frac{1}{1-\frac{\text{Var}_{|E_i\rangle}\left(T_1^y H_{\text{cut}}\right)}{\text{Var}_{|E_i\rangle}(H_{\text{cut}})}} = 4L\frac{1}{1-\text{CoR}_{|E_i\rangle}\left(H_{\text{cut}},H'_{\text{cut}}\right)}, \tag{B46}$$

where we dropped the positive term $\text{Var}_{|E_i\rangle}(SH^h_{2\text{-local}})$ in the first step. For $L \geq 2$ we even have

$$\mathcal{G}_{|E_i\rangle}\left(\mathcal{B}_{\text{Pauli}},\mathcal{B}_{L_x,L_y}\right) \geq 4L\frac{1+\frac{\text{Var}_{|E_i\rangle}\left(T_1^y H_{\text{cut}}\right)}{\text{Var}_{|E_i\rangle}(H_{\text{cut}})}}{1-\frac{\text{Var}_{|E_i\rangle}\left(T_1^y H_{\text{cut}}\right)}{\text{Var}_{|E_i\rangle}(H_{\text{cut}})}} = 4L\frac{1+\text{CoR}_{|E_i\rangle}\left(H_{\text{cut}},H'_{\text{cut}}\right)}{1-\text{CoR}_{|E_i\rangle}\left(H_{\text{cut}},H'_{\text{cut}}\right)}. \tag{B47}$$

At last, 2-local partitioning (compare Definition B2 c.) assumes minimal clusters of $L_x = 1$ and $L_y = 2$ and vice versa. A 2-local Hamiltonian has to be split into four parts, which do not commute in general, $H = H_1 + H_2 + H_3 + H_4$ (as defined in Definition B2 c.). For higher localities, the following discussion can be straightforwardly generalized. Since the four terms are connected by symmetry under translations and local rotations exchanging x and y, $H_1 = T_1^x H_2 T_{-1}^x = SH_3S = T_1^x SH_4 S T_{-1}^x$, their variances coincide. As a result, the sample improvement again becomes very simple,

$$\mathcal{G}_{|E_i\rangle}\left(\mathcal{B}_{\text{Pauli}},\mathcal{B}_{L_x,L_y}\right) \geq \left[\frac{\sqrt{\text{Var}_{|E_i\rangle}\left(H_{1\text{-local}}\right)}+\sqrt{\text{Var}_{|E_i\rangle}\left(H_{2\text{-local}}\right)}}{\sqrt{\text{Var}_{|E_i\rangle}\left(H_1\right)}+\sqrt{\text{Var}_{|E_i\rangle}\left(H_2\right)}+\sqrt{\text{Var}_{|E_i\rangle}\left(H_3\right)}+\sqrt{\text{Var}_{|E_i\rangle}\left(H_4\right)}}\right]^2 = \frac{1}{4}\frac{\text{Var}_{|E_i\rangle}(H_{2\text{-local}})}{\text{Var}_{|E_i\rangle}(H_1)}. \tag{B48}$$

We can use $\text{Var}(H_{2\text{-local}})$ from before and calculate the remaining variance of $H_1 = H_I + \frac{1}{4}H_{1\text{-local}} = H_I + \frac{1}{4}(H - H_{2\text{-local}})$, with the interactions $H_I = \sum_{\ell=1}^{n_x/2}\sum_{m=1}^{n_y} T_{2\ell}^x T_m^y V T_{-m}^y T_{-2\ell}^x$. The variance of H_1 can be calculated via $\text{Var}_{|E_i\rangle}(H_1) = \text{Var}_{|E_i\rangle}(O) + \frac{1}{16}\text{Var}_{|E_i\rangle}(H_{2\text{-local}}) - \frac{1}{4}(\text{CoV}_{|E_i\rangle}(O,H_{2\text{-local}}) + \text{CoV}_{|E_i\rangle}(H_{2\text{-local}},O))$. Using the same methods as before, we calculate

$$\text{Var}_{|E_i\rangle}(H_I) = n_y\frac{n_x}{2}\sum_{\ell=1}^{n_x/2}\sum_{m=1}^{n_y}\text{Var}_{|E_i\rangle}\left(T_{2\ell}^x T_m^y V\right), \tag{B49}$$

which is only the first term of C_2^x from Eq. (B31). We separate the factor $n/2$ and define the remainder of Eq. (B49) a new constant C_3. Similarly, the covariance reads

$$\text{CoV}_{|E_i\rangle}(O,H_{2\text{-local}}) = n_y\frac{n_x}{2}\sum_{\ell=1}^{n_x}\sum_{m=1}^{n_y}\text{CoV}_{|E_i\rangle}\left(V,T_\ell^x T_m^y V + T_\ell^x T_m^y SV\right) = \frac{n}{2}C_1 = \text{CoV}_{|E_i\rangle}(H_{2\text{-local}},O). \tag{B50}$$

Altogether we obtain

$$\mathcal{G}_{|E_i\rangle}\left(\mathcal{B}_{\text{Pauli}},\mathcal{B}_{L_x,L_y}\right) = \frac{1}{4}\frac{2nC_1}{\frac{n}{2}C_3+\frac{n}{8}C_1-\frac{n}{4}C_1} = \frac{C_1}{C_3-\frac{1}{4}C_1} \geq \frac{4}{3}, \tag{B51}$$

since $C_1 \geq C_3$.

APPENDIX C: SAMPLING IMPROVEMENTS UNDER NOISE (PROOFS OF SEC. IV)

1. Variance of isotropically perturbed states (Proof of Lemma 3)

We will use the flip operator defined by

$$\mathbb{F}\,|\psi_1\rangle\,|\psi_2\rangle := |\psi_2\rangle\,|\psi_1\rangle\,, \tag{C1}$$

the so-called swap-trick [e.g., see Eq. (7) of Ref. [50]],

$$\text{Tr}[AB] = \text{Tr}[(A\otimes B)\mathbb{F}], \tag{C2}$$

and that [e.g., see Eq. (163) of Ref. [50]],

$$\begin{aligned} \mathbb{E}\left[|\xi\rangle\right] = 0, \quad \mathbb{E}\left[|\xi\rangle\,|\xi\rangle\langle\xi|\right] = 0, \\ \mathbb{E}\left[|\xi\rangle\langle\xi|\right] = \mathbb{1}/d, \quad \mathbb{E}\left[|\xi\rangle\langle\xi|^{\otimes 2}\right] = \frac{1}{d(d+1)}(\mathbb{1}+\mathbb{F}). \end{aligned} \tag{C3}$$

Note that the last equality implies

$$\mathbb{E}\left[\langle\xi|O|\xi\rangle^2\right] = \mathbb{E}\left[\mathrm{Tr}[|\xi\rangle\langle\xi|^{\otimes 2}O^{\otimes 2}\mathbb{F}]\right] = \frac{1}{d(d+1)}\left(\mathrm{Tr}[O^{\otimes 2}\mathbb{F}^2] + \mathrm{Tr}[O^{\otimes 2}\mathbb{F}]\right) = \frac{\mathrm{Tr}[O]^2 + \|O\|_F^2}{d(d+1)}. \tag{C4}$$

Since $\mathrm{Var}(O) = \mathrm{Var}(O + \lambda\mathbb{1})$ we assume w.l.o.g. that $\mathrm{Tr}[O] = 0$. With these preliminaries, we then verify the claimed identity by a straightforward calculation,

$$\begin{aligned} \mathbb{E}\left[\mathrm{Var}_{|\tilde{\psi}\rangle}(O)\right] &= \mathbb{E}\left[\langle\tilde{\psi}|O^2|\tilde{\psi}\rangle - \langle\tilde{\psi}|O|\tilde{\psi}\rangle^2\right] = (1-\epsilon)\langle\psi|O^2|\psi\rangle + \epsilon\mathbb{E}\left[\langle\xi|O^2|\xi\rangle\right] - (1-\epsilon)^2\langle\psi|O|\psi\rangle^2 \\ &\quad - \epsilon^2\mathbb{E}\left[\langle\xi|O|\xi\rangle^2\right] - 4\epsilon(1-\epsilon)\mathbb{E}\left[\langle\psi|O|\xi\rangle\langle\xi|O|\psi\rangle\right] - 2\epsilon(1-\epsilon)\langle\psi|O|\psi\rangle\mathbb{E}[\langle\xi|O|\xi\rangle] \qquad \text{(C5)} \\ &= (1-\epsilon)\langle\psi|O^2|\psi\rangle + \epsilon\,\mathrm{Tr}\left[O^2\mathbb{E}|\xi\rangle\langle\xi|\right] - (1-\epsilon)^2\langle\psi|O|\psi\rangle^2 - \epsilon^2\mathbb{E}\left[\langle\xi|O|\xi\rangle^2\right] \\ &\quad - 4\epsilon(1-\epsilon)\,\langle\psi|O\mathbb{E}[|\xi\rangle\langle\xi|]O|\psi\rangle - 2\epsilon(1-\epsilon)\langle\psi|O|\psi\rangle\,\mathrm{Tr}\left[O\mathbb{E}|\xi\rangle\langle\xi|\right] \qquad \text{(C6)} \\ &= (1-\epsilon)\langle\psi|O^2|\psi\rangle - (1-\epsilon)^2\langle\psi|O|\psi\rangle^2 + \epsilon\|O\|_F^2/d - \frac{4\epsilon(1-\epsilon)}{d}\langle\psi|O^2|\psi\rangle \\ &\quad - \epsilon^2\frac{\mathrm{Tr}[O]^2 + \|O\|_F^2}{d(d+1)} - \frac{2\epsilon(1-\epsilon)}{d}\langle\psi|O|\psi\rangle\,\mathrm{Tr}\left[O\right] \qquad \text{(C7)} \\ &= (1-\epsilon)\langle\psi|O^2|\psi\rangle - (1-\epsilon)^2\langle\psi|O|\psi\rangle^2 + \epsilon\|O\|_F^2/d - \frac{4\epsilon(1-\epsilon)}{d}\langle\psi|O^2|\psi\rangle - \frac{\epsilon^2}{d(d+1)}\|O\|_F^2 \qquad \text{(C8)} \\ &= (1-\epsilon)\,\mathrm{Var}_{|\psi\rangle}(O) + \epsilon(1-\epsilon)\langle\psi|O|\psi\rangle^2 + \epsilon\|O\|_F^2/d - \frac{1}{d}\left(4\epsilon(1-\epsilon)\langle\psi|O|\psi\rangle^2 + \epsilon^2\|O\|^2_{\frac{F}{d+1}}\right). \qquad \text{(C9)} \end{aligned}$$

In Eq. (C9) the second and fourth, as well as the third and fifth term have the same structure apart from leading factors $4/d$ and ϵ/d. Thus, the last two terms are suppressed with the Hilbert space dimension $d = 2^n$, and we can simplify it to

$$\mathbb{E}\left[\mathrm{Var}_{|\tilde{\psi}\rangle}(O)\right] = (1-\epsilon)\,\mathrm{Var}_{|\psi\rangle}(O) + \epsilon(1-\epsilon)\langle\psi|O|\psi\rangle^2 + \epsilon\|O\|_F^2/d + \mathcal{O}\left(1/d\right). \tag{C10}$$

Further note that for O being a Hamiltonian H, a similar but simpler calculation tells us how the expectation value is gradually shifted from E_i to average energy $\bar{E}$, $|\tilde{\psi}\rangle = \sqrt{1-\epsilon}\,|E_i\rangle + \sqrt{\epsilon}\,|\xi\rangle$,

$$\mathbb{E}\left[\langle\tilde{\psi}|H|\tilde{\psi}\rangle\right] = (1-\epsilon)E_i + \epsilon\mathbb{E}\left[\langle\xi|H|\xi\rangle\right] = (1-\epsilon)E_i + \epsilon\frac{1}{d}\mathrm{Tr}(H) = (1-\epsilon)E_i + \epsilon\overline{E}. \tag{C11}$$

2. Relative sampling complexity improvement under global depolarization noise (Proof of Theorem 2)

As in the main text, we will concentrate here on bi-partitions $H = H_1 + H_2$, which apply for 1D and 2D geometric partitioning. At the same time, partitioning H into more parts, e.g., $H_2 = H_{2a} + H_{2b}$, will always only increase the overall sampling error due to Eq. (12), compare also Appendix A. Consequently, we will at worst underestimate the Pauli partitioning error.

As geometric partitioning distributes the couplings equally between H_1 and H_2, we have

$$\langle E_i|H^{(1)}_{L_x,L_y}|E_i\rangle = \langle E_i|H^{(2)}_{L_x,L_y}|E_i\rangle = \frac{E_i}{2} \quad \|H^{(1)}_{L_x,L_y}\|_F^2/d = \|H^{(2)}_{L_x,L_y}\|_F^2/d, \tag{C12}$$

and hence

$$\mathbb{E}\left[\mathrm{Var}_{|\tilde{E}_i\rangle}\left(H^{(1)}_{L_x,L_y}\right)\right] = \mathbb{E}\left[\mathrm{Var}_{|\tilde{E}_i\rangle}\left(H^{(2)}_{L_x,L_y}\right)\right]. \tag{C13}$$

I. Low randomness: The first statement to show is [Eq. (39)]

$$\epsilon \leq \frac{4\,\mathrm{Var}_{|E_i\rangle}\left(H^{(1)}_{L_x,L_y}\right)}{E_i^2 + \|H\|_F^2/d} \quad \Rightarrow \quad \overline{\mathcal{G}}_{|E_i\rangle}\left(\mathcal{B}^{\gamma\epsilon}_{\mathrm{Pauli}}, \mathcal{B}^{\epsilon}_{L_x,L_y}\right) \geq \frac{1}{2}\mathcal{G}_{|E_i\rangle}\left(\mathcal{B}_{\mathrm{Pauli}}, \mathcal{B}_{L_x,L_y}\right). \tag{C14}$$

From the requirement that the noise-free state-specific measurement should have lower than average case shot noise ($\|O\|_F^2/d \geq \mathrm{Var}_{|\psi\rangle}(O)$), if follows that:

$$\mathbb{E}\left[\mathrm{Var}_{|\tilde{\psi}(\epsilon)\rangle}(O)\right] = (1-\epsilon)\,\mathrm{Var}_{|\psi\rangle}(O) + \epsilon(1-\epsilon)\,\langle\psi|O|\psi\rangle^2 + \epsilon\|O\|_F^2/d \geq \mathrm{Var}_{|\psi\rangle}(O) + \epsilon\left[\|O\|_F^2/d - \mathrm{Var}_{|\psi\rangle}(O)\right] \tag{C15}$$

$$\geq \mathrm{Var}_{|\psi\rangle}(O) = \mathbb{E}\left[\mathrm{Var}_{|\tilde{\psi}(\epsilon=0)\rangle}(O)\right]. \tag{C16}$$

Next, we need to linearize Eq. (34) when plugging in Eq. (C15) from Lemma 3. The variance standard error of a partitioned observable $H = \sum H_b$ is given by

$$\left[\sum_{H_b\in\mathcal{B}}\sqrt{\mathbb{E}\left[\mathrm{Var}_{|E_i(\epsilon)\rangle}(H_b)\right]}\right]^2 = \left[\sum_{H_b\in\mathcal{B}}\sqrt{(1-\epsilon)\,\mathrm{Var}_{|E_i\rangle}(H_b) + \epsilon(1-\epsilon)\langle E_i|H_b|E_i\rangle^2 + \epsilon\|H_b\|_F^2/d}\right]^2 + \mathcal{O}(1/d). \tag{C17}$$

By labeling $A_b := (1-\epsilon)\,\mathrm{Var}_{|E_i\rangle}(H_b)$, $B_b := \epsilon(1-\epsilon)\langle E_i|H_b|E_i\rangle^2$ and $C_b := \epsilon\|H_b\|_F^2/d$, we simplify the notation. We will show

$$\left[\sum_b\sqrt{A_b+B_b}\right]^2 \geq \left[\sum_b\sqrt{A_b}\right]^2 + \left[\sum_b\sqrt{B_b}\right]^2, \tag{C18}$$

and thus by a nested argument

$$\left[\sum_b\sqrt{A_b+B_b+C_b}\right]^2 \geq \left[\sum_b\sqrt{A_b}\right]^2 + \left[\sum_b\sqrt{B_b+C_b}\right]^2 \geq \left[\sum_b\sqrt{A_b}\right]^2 + \left[\sum_b\sqrt{B_b}\right]^2 + \left[\sum_b\sqrt{C_b}\right]^2. \tag{C19}$$

Note that one is free to relabel A_b, B_b, and C_b to maximize the lower bound. We have

$$\left[\sum_b\sqrt{A_b+B_b}\right]^2 - \left[\sum_b\sqrt{A_b}\right]^2 - \left[\sum_b\sqrt{B_b}\right]^2 = \sum_{b\neq b'}\left[\sqrt{A_b+B_b}\sqrt{A_{b'}+B_{b'}} - \sqrt{A_bA_{b'}} - \sqrt{B_{b'}B_{b'}}\right]. \tag{C20}$$

Next, we will show that each term of the sum on the r.h.s. is larger or equal 0,

$$\left[\sqrt{A_bB_{b'}} - \sqrt{B_bA_{b'}}\right]^2 \geq 0, \tag{C21}$$

$$\Leftrightarrow A_bB_{b'} + B_bA_{b'} - 2\sqrt{A_bB_{b'}B_bA_{b'}} \geq 0, \tag{C22}$$

$$\Leftrightarrow A_bB_{b'} + B_bA_{b'} + A_bB_b + A_{b'}B_{b'} \geq 2\sqrt{A_bB_{b'}B_bA_{b'}} + A_bB_b + A_{b'}B_{b'}, \tag{C23}$$

$$\Leftrightarrow (A_b+B_b)(A_{b'}+B_{b'}) \geq (\sqrt{A_bA_{b'}} + \sqrt{B_bB_{b'}})^2. \tag{C24}$$

Thus Eq. (C17) can be lower bounded by

$$\left[\sum_{H_b\in\mathcal{B}}\sqrt{\mathbb{E}\left[\mathrm{Var}_{|E_i(\epsilon)\rangle}\,(H_b)\right]}\right]^2$$
$$\geq(1-\epsilon)\left[\sum_{H_b\in\mathcal{B}}\sqrt{\mathrm{Var}_{|E_i\rangle}(H_b)}\right]^2+\epsilon(1-\epsilon)\left[\sum_{H_b\in\mathcal{B}}\sqrt{\langle E_i|H_b|E_i\rangle^2}\right]^2+\epsilon\left[\sum_{H_b\in\mathcal{B}}\sqrt{\|H_b\|_F^2/d}\right]^2+\mathcal{O}(1/d). \tag{C25}$$

For the proofs of the shot-noise improvements under global depolarization noise, we will use Eq. (C25) extensively. Using further Eq. (C13) for the denominator, we get

$$\overline{\mathcal{G}}_{|E_i\rangle}\left(\mathcal{B}^{\gamma\epsilon}_{\mathrm{Pauli}},\mathcal{B}^{\epsilon}_{L_x,L_y}\right)\geq\overline{\mathcal{G}}_{|E_i\rangle}\left(\mathcal{B}^{0\epsilon}_{\mathrm{Pauli}},\mathcal{B}^{\epsilon}_{L_x,L_y}\right)=\frac{4\,\mathrm{Var}_{|E_i\rangle}\left(H^{(1)}_{\mathrm{Pauli}}\right)}{4(1-\epsilon)\,\mathrm{Var}_{|E_i\rangle}\left(H^{(1)}_{L_x,L_y}\right)+\epsilon(1-\epsilon)E_i^2+\epsilon(1+\beta)\|H\|_F^2/d}$$
$$\geq\frac{4\,\mathrm{Var}_{|E_i\rangle}\left(H^{(1)}_{\mathrm{Pauli}}\right)}{4(1-\epsilon)\,\mathrm{Var}_{|E_i\rangle}\left(H^{(1)}_{L_x,L_y}\right)+\epsilon(1-\epsilon)E_i^2+\epsilon\|H\|_F^2/d}, \tag{C26}$$

where $\beta_{L_x,L_y}:=\left|\|H^{(1)}_{L_x,L_y}\|_F^2/\|H\|_F^2-1/2\right|\leq 1$ gives the average asymmetry of squared Pauli weights across the partitions. We want to show that the r.h.s. of (C26) is bounded from below by $(1/2)\,\mathcal{G}_{|E_i\rangle}\left(\mathcal{B}_{\mathrm{Pauli}},\mathcal{B}_{L_x,L_y}\right)$. Hence it is sufficient to show that

$$4(1-\epsilon)\,\mathrm{Var}_{|E_i\rangle}\left(H^{(1)}_{L_x,L_y}\right)+\epsilon(1-\epsilon)E_i^2+\epsilon\|H\|_F^2/d\leq 2\cdot 4\,\mathrm{Var}_{|E_i\rangle}\left(H^{(1)}_{L_x,L_y}\right). \tag{C27}$$

When plugging in $\epsilon\leq(4\,\mathrm{Var}_{|E_i\rangle}\left(H^{(1)}_{L_x,L_y}\right))/(E_i^2+\|H\|_F^2/d)$, we obtain

$$4(1-\epsilon)\,\mathrm{Var}_{|E_i\rangle}\left(H^{(1)}_{L_x,L_y}\right)+\epsilon(1-\epsilon)E_i^2+\epsilon\|H\|_F^2/d\leq 4\,\mathrm{Var}_{|E_i\rangle}\left(H^{(1)}_{L_x,L_y}\right)\left[2-\epsilon\frac{2E_i^2+\|H\|_F^2/d}{E_i^2+\|H\|_F^2/d}\right] \tag{C28}$$
$$\leq 2\cdot 4\,\mathrm{Var}_{|E_i\rangle}\left(H^{(1)}_{L_x,L_y}\right). \tag{C29}$$

II. Intermediate randomness: Eq. (40) follows from

$$\overline{\mathcal{G}}_{|E_i\rangle}\left(\mathcal{B}^{\epsilon}_{L_x,L_y},\mathcal{B}^{\epsilon}_{H}\right)=\frac{4(1-\epsilon)\,\mathrm{Var}_{|E_i\rangle}\left(H^{(1)}_{L_x,L_y}\right)+\epsilon(1-\epsilon)E_i^2+4\epsilon\|H^{(1)}_{L_x,L_y}\|_F^2/d}{\epsilon(1-\epsilon)E_i^2+\epsilon\|H\|_F^2/d} \tag{C30}$$
$$\leq\frac{4\,\mathrm{Var}_{|E_i\rangle}\left(H^{(1)}_{L_x,L_y}\right)+\epsilon(1-\epsilon)E_i^2}{\epsilon(1-\epsilon)E_i^2+\epsilon\|H\|_F^2/d}\leq 1+\frac{4\,\mathrm{Var}_{|E_i\rangle}\left(H^{(1)}_{L_x,L_y}\right)}{\epsilon(1-\epsilon)E_i^2+\epsilon\|H\|_F^2/d}\overset{!}{\leq}2. \tag{C31}$$

This is true for

$$\epsilon\geq\frac{4\,\mathrm{Var}_{|E_i\rangle}\left(H^{(1)}_{L_x,L_y}\right)}{(1-\epsilon)E_i^2+\|H\|_F^2/d}\geq\frac{4\,\mathrm{Var}_{|E_i\rangle}\left(H^{(1)}_{L_x,L_y}\right)}{E_i^2+\|H\|_F^2/d}. \tag{C32}$$

By a similar calculation one shows Eq. (41),

$$\overline{\mathcal{G}}_{|E_i\rangle}\left(\mathcal{B}^{\gamma\epsilon}_{\text{Pauli}},\mathcal{B}^{\epsilon}_{H}\right)\geq\frac{4(1-\gamma\epsilon)\,\text{Var}_{|E_i\rangle}\left(H^{(1)}_{\text{Pauli}}\right)+\gamma\epsilon(1-\gamma\epsilon)(1+\alpha^2_{\text{Pauli}})E_i^2+\gamma\epsilon(1+\beta_{\text{Pauli}})\|H\|_F^2/d}{\epsilon(1-\epsilon)E_i^2+\epsilon\|H\|_F^2/d} \tag{C33}$$

$$\geq\frac{4(1-\epsilon)\,\text{Var}_{|E_i\rangle}\left(H^{(1)}_{\text{Pauli}}\right)}{\epsilon(1-\epsilon)E_i^2+\epsilon\|H\|_F^2/d}+\gamma\frac{(1-\epsilon)E_i^2+\|H\|_F^2/d}{(1-\epsilon)E_i^2+\|H\|_F^2/d}\overset{!}{\geq}2, \tag{C34}$$

where we used $\gamma\leq 1$, $\alpha_{\text{Pauli}}:=\left|\langle H^{(1)}_{\text{Pauli}}\rangle/E_i-1/2\right|\geq 0$ and $\beta_{\text{Pauli}}:=\left|\|H^{(1)}_{\text{Pauli}}\|_F^2/\|H\|_F^2-1/2\right|\geq 0$. Equation (C34) is true for

$$\epsilon\leq\frac{1}{2-\gamma}\frac{4\,\text{Var}_{|E_i\rangle}\left(H^{(1)}_{\text{Pauli}}\right)}{E_i^2+\frac{1}{1-\epsilon}\|H\|_F^2/d}\leq\frac{1}{2-\gamma}\frac{4\,\text{Var}_{|E_i\rangle}\left(H^{(1)}_{\text{Pauli}}\right)}{E_i^2+\|H\|_F^2/d}. \tag{C35}$$

III. Random dominated. Let $\mathcal{B}^{\gamma\epsilon}_{\text{bi–part.}}$ be a bi-partition of H ($H=H_1+H_2$) and $\overline{\mathcal{G}}_{|\tilde{E}_i\rangle}\left(\mathcal{B}^{\gamma\epsilon}_{\text{bi–part.}},\mathcal{B}^{\epsilon}_{H}\right)\leq C$,

$$\overline{\mathcal{G}}_{|\tilde{E}_i\rangle}\left(\mathcal{B}^{\gamma\epsilon}_{\text{bi–part.}},\mathcal{B}^{\epsilon}_{H}\right)\leq\frac{4(1-\gamma\epsilon)\,\text{Var}_{|E_i\rangle}\left(H^{(1)}_{\text{bi–part.}}\right)+\gamma\epsilon(1-\gamma\epsilon)2E_i^2+\gamma\epsilon 2\|H\|_F^2/d}{\epsilon(1-\epsilon)E_i^2+\epsilon\|H\|_F^2/d}\overset{!}{\leq}C, \tag{C36}$$

where we used $\text{Var}_{|E_i\rangle}\left(H^{(1)}_{\text{bi–part.}}\right)=\text{Var}_{|E_i\rangle}\left(H^{(2)}_{\text{bi–part.}}\right)$, $\alpha_{\text{bi–part.}}\leq 1$ and $\beta_{\text{bi–part.}}\leq 1$. Equation (C36) holds for

$$C-2\gamma\geq\frac{4\,\text{Var}_{|E_i\rangle}\left(H^{(1)}_{\text{bi–part.}}\right)}{\epsilon(1-\epsilon)E_i^2+\epsilon\|H\|_F^2/d}\geq\frac{4\,\text{Var}_{|E_i\rangle}\left(H^{(1)}_{\text{bi–part.}}\right)}{\epsilon E_i^2+\epsilon\|H\|_F^2/d}\Rightarrow\epsilon\geq\frac{1}{C-2\gamma}\frac{4\,\text{Var}_{|E_i\rangle}\left(H^{(1)}_{\text{bi–part.}}\right)}{E_i^2+\|H\|_F^2/d}. \tag{C37}$$

3. Criteria for sampling advantage of geometric partitioning under global depolarization (Proof of Corollary 2)

We have

$$\overline{\mathcal{G}}_{|E_i\rangle}\left(\mathcal{B}^{\gamma\epsilon}_{\text{Pauli}},\mathcal{B}^{\epsilon}_{L_x,L_y}\right)=\frac{\left[\sum_{H_b\in\mathcal{B}_{\text{Pauli}}}\sqrt{\mathbb{E}\left[\text{Var}_{|E_i(\gamma\epsilon)\rangle}\left(H_b\right)\right]}\right]^2}{4\mathbb{E}\left[\text{Var}_{|E_i(\epsilon)\rangle}\left(H^{(1)}_{L_x,L_y}\right)\right]} \tag{C38}$$

$$\geq\min\left[\frac{(1-\gamma\epsilon)}{(1-\epsilon)}\mathcal{G}_{|E_i\rangle}\left(\mathcal{B}_{\text{Pauli}},\mathcal{B}_{L_x,L_y}\right),\frac{\gamma\epsilon(1-\gamma\epsilon)}{\epsilon(1-\epsilon)}\frac{(1+\alpha^2_{\text{Pauli}})E_i^2}{E_i^2},\frac{\gamma\epsilon}{\epsilon}\left(\frac{\sum_{H_b\in\mathcal{B}_{\text{Pauli}}}\|H_b\|_F}{\sum_{H'_b\in\mathcal{B}_{L_x,L_y}}\|H'_b\|_F}\right)^2\right], \tag{C39}$$

where we first used Eq. (C25) on the numerator and then iteratively Eq. (44) ($(A+A')/(B+B')\geq\min(A/B,A'/B')$) to separate into variance-, E_i^2- and norm-type fractions. As $\mathcal{G}_{|E_i\rangle}\left(\mathcal{B}_{\text{Pauli}},\mathcal{B}_{L_x,L_y}\right)>1$ (Theorem 1), we get $\overline{\mathcal{G}}_{|E_i\rangle}\left(\mathcal{B}^{\gamma\epsilon}_{\text{Pauli}},\mathcal{B}^{\epsilon}_{L_x,L_y}\right)\geq 1$ for

$$\gamma(1+\alpha^2_{\text{Pauli}})\geq 1\quad\text{and}\quad\gamma\left(\frac{\sum_{H_b\in\mathcal{B}_{\text{Pauli}}}\|H_b\|_F}{\sum_{H'_b\in\mathcal{B}_{L_x,L_y}}\|H'_b\|_F}\right)^2\geq 1. \tag{C40}$$

Alternatively, we can also find a criterion for ϵ so that $\overline{\mathcal{G}}_{|E_i\rangle}\left(\mathcal{B}^{\gamma\epsilon}_{\text{Pauli}},\mathcal{B}^{\epsilon}_{L_x,L_y}\right)\geq 1$, similarly to Eq. (C26) we have

$$\overline{\mathcal{G}}_{|E_i\rangle}\left(\mathcal{B}^{\gamma\epsilon}_{\text{Pauli}},\mathcal{B}^{\epsilon}_{L_x,L_y}\right)\geq\frac{4\,\text{Var}_{|E_i\rangle}\left(H^{(1)}_{\text{Pauli}}\right)(1+\delta-\delta)}{4(1-\epsilon)\,\text{Var}_{|E_i\rangle}\left(H^{(1)}_{L_x,L_y}\right)+\epsilon(1-\epsilon)E_i^2+\epsilon\|H\|_F^2/d}. \tag{C41}$$

Next, we will use Eq. (44) again

$$\overline{\mathcal{G}}_{|E_i\rangle}\left(\mathcal{B}^{\gamma\epsilon}_{\text{Pauli}},\mathcal{B}^{\epsilon}_{L_x,L_y}\right)\geq\min\left(\frac{\delta\mathcal{G}_{|E_i\rangle}\left(\mathcal{B}_{\text{Pauli}},\mathcal{B}_{L_x,L_y}\right)}{1-\epsilon},\frac{4\,\text{Var}_{|E_i\rangle}\left(H^{(1)}_{\text{Pauli}}\right)(1-\delta)}{\epsilon(1-\epsilon)E_i^2+\epsilon\|H\|_F^2/d}\right)\overset{!}{\geq}1, \tag{C42}$$

if we set $\delta=1/\mathcal{G}_{|E_i\rangle}\left(\mathcal{B}_{\text{Pauli}},\mathcal{B}_{L_x,L_y}\right)$ then the first term will always be larger than 1 as $\frac{1}{1-\epsilon}\geq 1$, while the second term will be larger 1 for

$$\frac{1}{\epsilon}\frac{4\,\text{Var}_{|E_i\rangle}\left(H^{(1)}_{\text{Pauli}}\right)(1-\delta)}{(1-\epsilon)E_i^2+\|H\|_F^2/d}\geq\frac{1-\delta}{\epsilon}\frac{4\,\text{Var}_{|E_i\rangle}\left(H^{(1)}_{\text{Pauli}}\right)}{E_i^2+\|H\|_F^2/d}\overset{!}{\geq}1\quad\Rightarrow\quad\epsilon\leq\left(1-\frac{1}{\mathcal{G}_{|E_i\rangle}\left(\mathcal{B}_{\text{Pauli}},\mathcal{B}_{L_x,L_y}\right)}\right)\frac{4\,\text{Var}_{|E_i\rangle}\left(H^{(1)}_{\text{Pauli}}\right)}{E_i^2+\|H\|_F^2/d}. \tag{C43}$$

4. Gate budget for noisy geometric partitioning (Proof of Corollary 3)

Here, we use Theorem 2 to derive a criterion that allows a sample improvement of at least a factor G,

$$\overline{\mathcal{G}}_{|E_i\rangle}\left(\mathcal{B}^{\gamma\epsilon}_{\text{Pauli}},\mathcal{B}^{\epsilon}_{L_x,L_y}\right)\geq G. \tag{C44}$$

For this, we replace the bound in Theorem 2 in Eq. (C42) [main text Eq. (48)] "≥ 1" with "$\geq G$". This gives

$$\epsilon\leq\left(1-\frac{G}{\mathcal{G}_{|E_i\rangle}\left(\mathcal{B}_{\text{Pauli}},\mathcal{B}_{L_x,L_y}\right)}\right)\epsilon_{II\leftrightarrow III}. \tag{C45}$$

Further, we keep Eq. (C31) [main text Eq. (40)] and replace the "2" at the r.h.s. with $2G$. This gives the alternative inequality

$$\epsilon\leq\frac{1}{2G-\gamma}\epsilon_{II\leftrightarrow III}. \tag{C46}$$

By choosing the maximum of both, $c_\epsilon=\max(1/(2G-\gamma),\left[1-G/\mathcal{G}_{|E_i\rangle}\left(\mathcal{B}_{\text{Pauli}},\mathcal{B}_{L_x,L_y}\right)\right])$, we get

$$1-(1-\epsilon_{\text{gate}})^{\mathcal{N}_{2q}(U_{|E_i\rangle})+\mathcal{N}_{2q}\left(U_{L_x,L_y}\right)}\approx\epsilon\leq c_\epsilon\epsilon_{II\leftrightarrow III}\;\Leftrightarrow\quad\mathcal{N}_{2q}(U_{|E_i\rangle})+\mathcal{N}_{2q}\left(U_{L_x,L_y}\right)\lessapprox\frac{\log\left(1-c_\epsilon\epsilon_{II\leftrightarrow III}\right)}{\log(1-\epsilon_{\text{gate}})}. \tag{C47}$$

APPENDIX D: ADDITIONAL MATERIAL FOR NUMERICAL TESTS

1. Proof of Lemma 2 "Disordered phase" $h\gg J$

We normalize the Hamiltonian to obtain (this does not change the variance ratio)

$$H=-\sum_i X_i-\left(\frac{J}{h}\right)\sum_{\langle i,j\rangle}Z_iZ_j=-\sum_i X_i-\lambda\sum_{\langle i,j\rangle}Z_iZ_j=H_0+\lambda V. \tag{D1}$$

The ground state $|E_0\rangle$ is given up to second order in λ by

$$|E_0^{(0)}\rangle=|+\rangle^{\otimes n}, \tag{D2}$$

$$|E_0^{(1)}\rangle=\sum_{|b\rangle\neq|+\rangle^{\otimes n}}\frac{\langle b|\sum_{\langle i,j\rangle}Z_iZ_j|+^{\otimes n}\rangle}{-n-(n-2\cdot 2)}|b\rangle=\frac{1}{4}\sum_{\langle i,j\rangle}Z_iZ_j\,|+\rangle^{\otimes n}, \tag{D3}$$

$$|E_0^{(2)}\rangle = \sum_{|b_1\rangle\neq|+\rangle^{\otimes n}} \sum_{|b_2\rangle\neq|+\rangle^{\otimes n}} |b_1\rangle \frac{\langle b_1|\sum_{\langle i,j\rangle} Z_iZ_j|b_2\rangle\langle b_2|\sum_{\langle i,j\rangle} Z_iZ_j|+^{\otimes n}\rangle}{\left(E^{(0)}_{|+\rangle^{\otimes n}} - E^{(0)}_{|b_1\rangle}\right)\left(E^{(0)}_{|+\rangle^{\otimes n}} - E^{(0)}_{|b_2\rangle}\right)} - \frac{1}{2}|+\rangle^{\otimes n}\frac{(2n)}{4^2} \tag{D4}$$

$$= \left[\frac{2}{4\cdot 8}\sum_{\substack{\langle i,j\rangle,\langle k,\ell\rangle \\ i,j\neq k,\ell}} (Z_iZ_j)(Z_kZ_\ell) + \frac{4}{16}\sum_{\langle\langle i,j\rangle\rangle_{\text{diagonal}}} Z_iZ_j + \frac{2}{16}\sum_{\langle\langle i,j\rangle\rangle_{\text{axial}}} Z_iZ_j - \frac{n}{16}\mathbb{1}\right]|+\rangle^{\otimes n}, \tag{D5}$$

$$|E_0\rangle = \left[\mathbb{1} + \frac{\lambda}{4}\sum_{\langle i,j\rangle} Z_iZ_j + \frac{\lambda^2}{16}\left(\sum_{\substack{\langle i,j\rangle,\langle k,\ell\rangle \\ i,j\neq k,\ell}} (Z_iZ_j)(Z_kZ_\ell) + 4\sum_{\langle\langle i,j\rangle\rangle_{\text{diagonal}}} Z_iZ_j + 2\sum_{\langle\langle i,j\rangle\rangle_{\text{axial}}} Z_iZ_j - n\mathbb{1}\right)\right]|+\rangle^{\otimes n} + \mathcal{O}(\lambda^3) \tag{D6}$$

$$= P_X^{(2)}|+\rangle^{\otimes n} + \mathcal{O}(\lambda^3). \tag{D7}$$

Note that $\sum_{\langle\langle i,j\rangle\rangle_{\text{axial}}}$ obtains a factor 2 if $n_x, n_y = 4$ and does not exist for $n_x, n_y = 3$. Let us start with calculating the variance of H_0 up to second order,

$$\text{Var}(H_0) = \langle (H_0)^2\rangle - \langle H_0\rangle^2, \tag{D8}$$

$$\langle H_0\rangle = \langle +|^{\otimes n}\left[\mathbb{1} + \frac{\lambda}{4}\sum_{\langle i,j\rangle} Z_iZ_j + \frac{\lambda^2}{16}\left(\sum_{\substack{\langle i,j\rangle,\langle k,\ell\rangle \\ i,j\neq k,\ell}} (Z_iZ_j)(Z_kZ_\ell) + 4\sum_{\substack{\langle\langle i,j\rangle\rangle \\ \text{diagonal}}} Z_iZ_j + 2\sum_{\substack{\langle\langle i,j\rangle\rangle \\ \text{axial}}} Z_iZ_j - n\mathbb{1}\right)\right] \tag{D9}$$

$$\cdot \sum_i X_i\left[\mathbb{1} + \frac{\lambda}{4}\sum_{\langle i,j\rangle} Z_iZ_j + \frac{\lambda^2}{16}\left(\sum_{\substack{\langle i,j\rangle,\langle k,\ell\rangle \\ i,j\neq k,\ell}} (Z_iZ_j)(Z_kZ_\ell) + 4\sum_{\substack{\langle\langle i,j\rangle\rangle \\ \text{diagonal}}} Z_iZ_j + 2\sum_{\substack{\langle\langle i,j\rangle\rangle \\ \text{axial}}} Z_iZ_j - n\mathbb{1}\right)\right]|+\rangle^{\otimes n} + \mathcal{O}(\lambda^3) \tag{D10}$$

$$= n + \lambda^2\left(\frac{1}{4^2}(2n(n-2-2)) - 2\frac{n^2}{16}\right) + \mathcal{O}(\lambda^3) = n - \frac{\lambda^2}{2}n + \mathcal{O}(\lambda^3), \tag{D11}$$

$$\langle (H_0)^2\rangle = \langle +|^{\otimes n}\left[\mathbb{1} + \frac{\lambda}{4}\sum_{\langle i,j\rangle} Z_iZ_j + \frac{\lambda^2}{16}\left(\sum_{\substack{\langle i,j\rangle,\langle k,\ell\rangle \\ i,j\neq k,\ell}} (Z_iZ_j)(Z_kZ_\ell) + 4\sum_{\substack{\langle\langle i,j\rangle\rangle \\ \text{diagonal}}} Z_iZ_j + 2\sum_{\substack{\langle\langle i,j\rangle\rangle \\ \text{axial}}} Z_iZ_j - n\mathbb{1}\right)\right]\left[n\mathbb{1} + \sum_{i\neq j} X_iX_j\right] \tag{D12}$$

$$\cdot \left[\mathbb{1} + \frac{\lambda}{4}\sum_{\langle i,j\rangle} Z_iZ_j + \frac{\lambda^2}{16}\left(\sum_{\substack{\langle i,j\rangle,\langle k,\ell\rangle \\ i,j\neq k,\ell}} (Z_iZ_j)(Z_kZ_\ell) + 4\sum_{\substack{\langle\langle i,j\rangle\rangle \\ \text{diagonal}}} Z_iZ_j + 2\sum_{\substack{\langle\langle i,j\rangle\rangle \\ \text{axial}}} Z_iZ_j - n\mathbb{1}\right)\right]|+\rangle^{\otimes n} + \mathcal{O}(\lambda^3) \tag{D13}$$

$$= n^2 + \lambda^2\left(\frac{2n[n + (n-2)(n-3) + 2 - 4(n-2)]}{16} - \frac{2n^3}{16}\right) + \mathcal{O}(\lambda^3) = n^2 + \frac{\lambda^2}{16}\left(-16n^2 + 32n\right) + \mathcal{O}(\lambda^3), \tag{D14}$$

$$\text{Var}(H_0) = \langle (H_0)^2\rangle - \langle H_0\rangle^2 = n^2 + \lambda^2\left(2n - n^2\right) - \left[n^2 - n^2\lambda^2\right] + \mathcal{O}(\lambda^3) = 2n\lambda^2 + \mathcal{O}(\lambda^3). \tag{D15}$$

Next let us calculate (for $L > 1$)

$$\text{Var}(H_l - H_r) = \langle (H_l - H_r)^2\rangle - \langle H_l - H_r\rangle^2 = 2\langle (H_l)^2\rangle - 2\langle H_lH_r\rangle \tag{D16}$$

$$H_l = \lambda \sum_{i=1}^{n_x/L} \sum_{j=1}^{n_y} Z_{Li,j} Z_{Li+1,j} \quad H_r = \lambda \sum_{i=1}^{n_x/L} \sum_{j=1}^{n_y} Z_{Li+1,j} Z_{Li+2,j}\,. \tag{D17}$$

This calculation is also true for $L = 3$ as there the axial next-nearest neighbors would be neighbors, but the term in the expansion does not exist then,

$$(\langle E_0|\, H_l)\,(H_l\,|E_0\rangle) \tag{D18}$$

$$= \langle +|^{\otimes n} \left[\mathbb{1} + \frac{\lambda}{4} \sum_{\langle i,j\rangle} Z_i Z_j + \frac{\lambda^2}{16} \left(\sum_{\substack{\langle i,j\rangle,\langle k,\ell\rangle \\ i,j\neq k,\ell}} (Z_i Z_j)(Z_k Z_\ell) + 4 \sum_{\substack{\langle\langle i,j\rangle\rangle \\ \text{diagonal}}} Z_i Z_j + 2 \sum_{\substack{\langle\langle i,j\rangle\rangle \\ \text{axial}}} Z_i Z_j - n\mathbb{1} \right) \right] \tag{D19}$$

$$\cdot \lambda^2 \left[\frac{n}{L}\mathbb{1} + \sum_{i=1}^{n_x/L} \sum_{j=1}^{n_y} \sum_{k=1}^{n_x/L} \sum_{\ell=1,(i,j)\neq(k,\ell)}^{n_y} Z_{Li,j} Z_{Li+1,j} Z_{Lk,\ell} Z_{Lk+1,\ell} \right] \tag{D20}$$

$$\cdot \left[\mathbb{1} + \frac{\lambda}{4} \sum_{\langle i,j\rangle} Z_i Z_j + \frac{\lambda^2}{16} \left(\sum_{\substack{\langle i,j\rangle,\langle k,\ell\rangle \\ i,j\neq k,\ell}} (Z_i Z_j)(Z_k Z_\ell) + 4 \sum_{\substack{\langle\langle i,j\rangle\rangle \\ \text{diagonal}}} Z_i Z_j + 2 \sum_{\substack{\langle\langle i,j\rangle\rangle \\ \text{axial}}} Z_i Z_j - n\mathbb{1} \right) \right] |+\rangle^{\otimes n} + \mathcal{O}(\lambda^3) \tag{D21}$$

$$= \frac{n}{L}\lambda^2 + \frac{\lambda^4}{16} \left(\frac{n}{L} 2 \cdot 2n + 2 \cdot \frac{n}{L} \cdot \frac{n-1}{L} - 2 \cdot \frac{n^2}{L} + 2 \cdot 2 \cdot \frac{n}{L} \cdot \frac{n-1}{L} \right) + \mathcal{O}(\lambda^5) \tag{D22}$$

$$= \frac{n}{L}\lambda^2 + \frac{\lambda^4}{16}\frac{n^2}{L} \left(4 + \frac{2}{L} - \frac{2}{nL} - 2 + \frac{4}{L} - \frac{4}{nL} \right) + \mathcal{O}(\lambda^5) = \frac{n}{L}\lambda^2 + \frac{\lambda^4}{16}\frac{n^2}{L} \left(2 + \frac{6}{L}\left(1 - \frac{1}{n}\right) \right) + \mathcal{O}(\lambda^5), \tag{D23}$$

$$(\langle E_0|\, H_l)\,(H_r\,|E_0\rangle) \tag{D24}$$

$$= \langle +|^{\otimes n} \left[\mathbb{1} + \frac{\lambda}{4} \sum_{\langle i,j\rangle} Z_i Z_j + \frac{\lambda^2}{16} \left(\sum_{\substack{\langle i,j\rangle,\langle k,\ell\rangle \\ i,j\neq k,\ell}} (Z_i Z_j)(Z_k Z_\ell) + 4 \sum_{\substack{\langle\langle i,j\rangle\rangle \\ \text{diagonal}}} Z_i Z_j + 2 \sum_{\substack{\langle\langle i,j\rangle\rangle \\ \text{axial}}} Z_i Z_j - n\mathbb{1} \right) \right] \tag{D25}$$

$$\cdot \left[\lambda \sum_{i=1}^{n_x/L} \sum_{j=1}^{n_y} Z_{Li,j} Z_{Li+1,j} \right] \left[\lambda \sum_{i=1}^{n_x/L} \sum_{j=1}^{n_y} Z_{Li+1,j} Z_{Li+2,j} \right] \tag{D26}$$

$$\cdot \left[\mathbb{1} + \frac{\lambda}{4} \sum_{\langle i,j\rangle} Z_i Z_j + \frac{\lambda^2}{16} \left(\sum_{\substack{\langle i,j\rangle,\langle k,\ell\rangle \\ i,j\neq k,\ell}} (Z_i Z_j)(Z_k Z_\ell) + 4 \sum_{\substack{\langle\langle i,j\rangle\rangle \\ \text{diagonal}}} Z_i Z_j + 2 \sum_{\substack{\langle\langle i,j\rangle\rangle \\ \text{axial}}} Z_i Z_j - n\mathbb{1} \right) \right] |+\rangle^{\otimes n} + \mathcal{O}(\lambda^3\lambda^2) \tag{D27}$$

$$= \mathcal{O}(\lambda^4). \tag{D28}$$

Thus we have

$$\mathrm{Var}(H_0) = 2n\lambda^2 + \mathcal{O}(\lambda^3), \tag{D29}$$

$$\mathrm{Var}(H_l - H_r) = 2 \cdot \frac{n}{L}\lambda^2 + \mathcal{O}(\lambda^4), \tag{D30}$$

$$\frac{\mathrm{Var}(H_0)}{\mathrm{Var}(\frac{1}{2}(H - H_l + H_r))} = 4\frac{\mathrm{Var}(H_1)}{\mathrm{Var}(H_l - H_r)} = 4\frac{2n\lambda^2}{2 \cdot \frac{n}{L}\lambda^2} + \mathcal{O}(\lambda) = 4L + \mathcal{O}(\lambda). \tag{D31}$$

2. "Ordereds phase" $h \ll J$

We normalize the Hamiltonian to obtain (this does not change the variance ratio)

$$H = -\sum_{\langle i,j\rangle} Z_i Z_j - \left(\frac{h}{J}\right)\sum_i X_i = -\sum_{\langle i,j\rangle} Z_i Z_j - \lambda \sum_i X_i = H_0 + \lambda V. \tag{D32}$$

Then we can calculate the perturbed state order by order

$$\frac{\mathrm{Var}(H_2)}{\mathrm{Var}(\frac{1}{2}(H - H_l + H_r))} = 4\frac{\mathrm{Var}(H_1)}{\mathrm{Var}(H_l - H_r)}. \tag{D33}$$

We now want to calculate

$$|E_0^{(0)}\rangle = |0\rangle^{\otimes n}, \tag{D34}$$

$$|E_0^{(1)}\rangle = \sum_{|b\rangle\neq|0\rangle^{\otimes n}} \frac{\langle b|\sum_i X_i|0^{\otimes n}\rangle}{-2n - E_k^{(0)}}|b\rangle = \frac{1}{8}\sum_i X_i|0\rangle^{\otimes n}, \tag{D35}$$

$$|E_0^{(2)}\rangle = \sum_{|b_1\rangle\neq|0\rangle^{\otimes n}}\sum_{|b_2\rangle\neq|0\rangle^{\otimes n}} |b_1\rangle \frac{\langle b_1|\sum_i X_i|b_2\rangle\langle b_2|\sum_i X_i|0^{\otimes n}\rangle}{\left(E_n^{(0)} - E_k^{(0)}\right)\left(E_n^{(0)} - E_\ell^{(0)}\right)} \tag{D36}$$

$$- \sum_{k\neq n}\left|k^{(0)}\right\rangle \frac{\left\langle k^{(0)}\right| V \left|n^{(0)}\right\rangle\langle 0^{\otimes n}|\sum_i X_i|0^{\otimes n}\rangle}{\left(E_n^{(0)} - E_k^{(0)}\right)^2} - \frac{1}{2}|0\rangle^{\otimes n}\frac{n}{8^2} \tag{D37}$$

$$= \left(\frac{2}{8\cdot 2\cdot 6}\sum_{\langle i,j\rangle} X_iX_j + \frac{1}{8\cdot 2\cdot 8}\sum_{i\neq j\setminus\langle i,j\rangle} X_iX_j - \frac{1}{128}n\mathbb{1}\right)|0\rangle^{\otimes n}, \tag{D38}$$

$$|E_0\rangle = \left[\mathbb{1} + \frac{\lambda}{8}\sum_i X_i + \frac{\lambda^2}{16}\left(\frac{1}{3}\sum_{\langle i,j\rangle} X_iX_j + \frac{1}{8}\sum_{i\neq j\setminus\langle i,j\rangle} X_iX_j - \frac{1}{8}n\mathbb{1}\right)\right]|0\rangle^{\otimes n} + \mathcal{O}(\lambda^3). \tag{D39}$$

Note here that the $\sum_{\langle i,j\rangle}$ receives a factor 2 as then there is no double counting of lattice sides any more, while $\sum_{i\neq j\setminus\langle i,j\rangle}$ still has every term twice.

Further note that we defined $H_r = T_{-1}H_lT_1$. In the $L = 1$ case we use instead $H_r = SH_lS$. Further, we could have taken $H_r = T_{-\lceil L/2\rceil}H_lT_{\lceil L/2\rceil}$. Indeed, it turns out that the last definition performs a factor 2 worse than the first one in the perturbative limit.

Let us start with calculating the variance of $H_1 = \lambda\sum_i X_i$ up to second order,

$$\mathrm{Var}(H_1) = \langle (H_1)^2\rangle - \langle H_1\rangle^2, \tag{D40}$$

$$\langle (H_1)^2\rangle = \lambda^2 n + \lambda^2\langle 0|^{\otimes n}\left[\mathbb{1} + \frac{\lambda}{8}\sum_i X_i + \frac{\lambda^2}{16}\left(\frac{1}{3}\sum_{\langle i,j\rangle} X_iX_j + \frac{1}{8}\sum_{i\neq j\setminus\langle i,j\rangle} X_iX_j - \frac{1}{8}n\mathbb{1}\right)\right]\left[\sum_{i\neq j} X_iX_j\right] \tag{D41}$$

$$\cdot\left[\mathbb{1} + \frac{\lambda}{8}\sum_i X_i + \frac{\lambda^2}{16}\left(\frac{1}{3}\sum_{\langle i,j\rangle} X_iX_j + \frac{1}{8}\sum_{i\neq j\setminus\langle i,j\rangle} X_iX_j - \frac{1}{8}n\mathbb{1}\right)\right]|0\rangle^{\otimes n} + \mathcal{O}(\lambda^3\lambda^2) \tag{D42}$$

$$= \lambda^2 n + \lambda^4\left(\frac{n\cdot 2(n-1)}{8^2} + 2\frac{2n\cdot 2}{3\cdot 16} + 2\frac{[n(n-1) - 2n]\cdot 2}{8\cdot 16}\right) + \mathcal{O}(\lambda^3\lambda^2) \tag{D43}$$

$$= \lambda^2 n + \lambda^4\left(n^2\left(\frac{1}{32} + \frac{1}{32}\right) + n\left(-\frac{1}{32} + \frac{1}{6} - \frac{3}{32}\right)\right) + \mathcal{O}(\lambda^5) = \lambda^2 n + \frac{\lambda^4}{16}\left(n^2 - \frac{4}{3}n\right) + \mathcal{O}(\lambda^5), \tag{D44}$$

$$\langle H_1\rangle = \lambda\,\langle 0|^{\otimes n}\left[\mathbb{1}+\frac{\lambda}{8}\sum_i X_i+\frac{\lambda^2}{16}\left(\frac{1}{3}\sum_{\langle i,j\rangle}X_iX_j+\frac{1}{8}\sum_{i\neq j\setminus\langle i,j\rangle}X_iX_j-\frac{1}{8}n\mathbb{1}\right)\right]\left[\sum_i X_i\right] \tag{D45}$$

$$\cdot\left[\mathbb{1}+\frac{\lambda}{8}\sum_i X_i+\frac{\lambda^2}{16}\left(\frac{1}{3}\sum_{\langle i,j\rangle}X_iX_j+\frac{1}{8}\sum_{i\neq j\setminus\langle i,j\rangle}X_iX_j-\frac{1}{8}n\mathbb{1}\right)\right]|0\rangle^{\otimes n}+\mathcal{O}(\lambda^3\lambda^1) \tag{D46}$$

$$=\lambda^2\left(\frac{2}{8}n\right)+\mathcal{O}(\lambda^4), \tag{D47}$$

$$\mathrm{Var}(H_1)=n\lambda^2-\frac{1}{12}n\lambda^4+\mathcal{O}(\lambda^5). \tag{D48}$$

Let us calculate the variance of H_l-H_r up to second order. Let us consider a general H_r which is translated with respect to H_l by $i_r\le L/2$,

$$\mathrm{Var}\left(\frac{1}{2}(H-H_l+H_r)\right)=\frac{1}{4}\mathrm{Var}(H_l-H_r)=\frac{1}{2}\left(\langle (H_l)^2\rangle-\langle H_lH_r\rangle\right), \tag{D49}$$

$$H_l=\sum_{i=1}^{n_x/L}\sum_{j=1}^{n_y}Z_{Li,j}Z_{Li+1,j}\qquad H_r=\sum_{i=1}^{n_x/L}\sum_{j=1}^{n_y}Z_{Li+i_r,j}Z_{Li+i_r+1,j}, \tag{D50}$$

$$\langle (H_l)^2\rangle=\frac{n}{L}+\langle 0|^{\otimes n}\left[\mathbb{1}+\frac{\lambda}{8}\sum_i X_i+\frac{\lambda^2}{16}\left(\frac{1}{3}\sum_{\langle i,j\rangle}X_iX_j+\frac{1}{8}\sum_{i\neq j\setminus\langle i,j\rangle}X_iX_j-\frac{1}{8}n\mathbb{1}\right)\right] \tag{D51}$$

$$\cdot\left[\sum_{i,k=1}^{n_x/L}\sum_{\substack{j,\ell=1\\(i,j)\neq(k,\ell)}}^{n_y}(Z_{Li,j}Z_{Li+1,j})(Z_{Lk,\ell}Z_{Lk+1,\ell})\right] \tag{D52}$$

$$\cdot\left[\mathbb{1}+\frac{\lambda}{8}\sum_i X_i+\frac{\lambda^2}{16}\left(\frac{1}{3}\sum_{\langle i,j\rangle}X_iX_j+\frac{1}{8}\sum_{i\neq j\setminus\langle i,j\rangle}X_iX_j-\frac{1}{8}n\mathbb{1}\right)\right]|0\rangle^{\otimes n}+\mathcal{O}(\lambda^4) \tag{D53}$$

$$=\left(\frac{n}{L}\right)^2+\frac{\lambda^2}{64}\left(\frac{n}{L}\left(\frac{n}{L}-1\right)(n-2\cdot 4)-2\frac{n}{2}\frac{n}{L}\left(\frac{n}{L}-1\right)\right)+\mathcal{O}(\lambda^4)=\left(\frac{n}{L}\right)^2-\frac{\lambda^2}{8}\frac{n}{L}\left(\frac{n}{L}-1\right)$$
$$+\mathcal{O}(\lambda^4), \tag{D54}$$

$$\langle H_lH_r\rangle=\langle 0|^{\otimes n}\left[\mathbb{1}+\frac{\lambda}{8}\sum_i X_i+\frac{\lambda^2}{16}\left(\frac{1}{3}\sum_{\langle i,j\rangle}X_iX_j+\frac{1}{8}\sum_{i\neq j\setminus\langle i,j\rangle}X_iX_j-\frac{1}{8}n\mathbb{1}\right)\right] \tag{D55}$$

$$\cdot\left[\delta_{L,2}\delta_{i_r,1}\sum_{i=1}^{n_x/L}\sum_{j=1}^{n_y}Z_{Li-1,j}Z_{Li+1,j}+\delta_{i_r,1}\sum_{i=1}^{n_x/L}\sum_{j=1}^{n_y}Z_{Li,j}Z_{Li+2,j}\right. \tag{D56}$$

$$\left.+\sum_{i,k=1}^{n_x/L}\sum_{\substack{j,\ell=1\\(Li,j)\neq(Lk+i_r+\{0\pm1\},\ell)}}^{n_y}(Z_{Li,j}Z_{Li+1,j})(Z_{Lk+i_r,\ell}Z_{Lk+i_r+1,\ell})\right] \tag{D57}$$

$$\cdot\left[\mathbb{1}+\frac{\lambda}{8}\sum_i X_i+\frac{\lambda^2}{16}\left(\frac{1}{3}\sum_{\langle i,j\rangle}X_iX_j+\frac{1}{8}\sum_{i\neq j\setminus\langle i,j\rangle}X_iX_j-\frac{1}{8}n\mathbb{1}\right)\right]|0\rangle^{\otimes n}+\mathcal{O}(\lambda^3) \tag{D58}$$

$$= \left(\frac{n}{L}\right)^2 + \frac{\lambda^2}{64}\left((\delta_{L,2}+\delta_{i_r,1})n\cdot\frac{n-2\cdot 2}{L} + (n-2\cdot 4)\frac{n}{L}\left(\frac{n}{L}-\delta_{L,2}-\delta_{i_r,1}\right) - 2\frac{n^3}{2L^2}\right) + O(\lambda^4) \tag{D59}$$

$$= \left(\frac{n}{L}\right)^2 + \frac{\lambda^2}{16}\frac{n^2}{L}\left(\left(\frac{n-8}{L} + (\delta_{L,2}+\delta_{i_r,1})\frac{4}{n}\right) - \frac{n}{4L}\right) + O(\lambda^4). \tag{D60}$$

Note that $L=2$ implies $i_r = 1$. We have $O(\lambda^4)$, not only $O(\lambda^3)$ as the third perturbation order contains single and triple excitations and thus the product of zeroth and third order terms vanish as well. The third sum has $n^2/L^2 - (\delta_{L,2}+\delta_{i_r,1})n/L$ elements. Whenever XZX act on the same qubits, we obtain $-Z$ and we obtain $+Z$ if they act on different qubits. For $L=2$, we have

$$H_l = \sum_{i=1}^{n_x/L}\sum_{j=1}^{n_y} Z_{2i,j}Z_{2i+1,j} \quad H_r = \sum_{i=1}^{n_x/L}\sum_{j=1}^{n_y} Z_{2i+1,j}Z_{2i+2,j}, \tag{D61}$$

$$H_lH_r = \sum_{i=1}^{n_x/L}\sum_{j=1}^{n_y}\sum_{k=1}^{n_x/L}\sum_{l=1}^{n_y} Z_{2i,j}Z_{2i+1,j}Z_{2k+1,l}Z_{2k+2,l}. \tag{D62}$$

So

$$\langle H_lH_r\rangle = \langle 0|^{\otimes n}\left[\mathbb{1} + \frac{\lambda}{8}\sum_i X_i + \frac{\lambda^2}{16}\left(\frac{1}{3}\sum_{\langle i,j\rangle}X_iX_j + \frac{1}{8}\sum_{i\neq j\setminus\langle i,j\rangle}X_iX_j - \frac{1}{8}n\mathbb{1}\right)\right] \tag{D63}$$

$$\cdot\sum_{i=1}^{n_x/L}\sum_{j=1}^{n_y}\sum_{k=1}^{n_x/L}\sum_{l=1}^{n_y} Z_{2i,j}Z_{2i+1,j}Z_{2k+1,l}Z_{2k+2,l} \tag{D64}$$

$$\cdot\left[\mathbb{1} + \frac{\lambda}{8}\sum_i X_i + \frac{\lambda^2}{16}\left(\frac{1}{3}\sum_{\langle i,j\rangle}X_iX_j + \frac{1}{8}\sum_{i\neq j\setminus\langle i,j\rangle}X_iX_j - \frac{1}{8}n\mathbb{1}\right)\right]|0\rangle^{\otimes n} + \mathcal{O}(\lambda^3) \tag{D65}$$

$$= \left(\frac{n}{L}\right)^2 + \frac{\lambda^2}{64}\left(2\frac{n}{L}(n-4) + \left(\left(\frac{n}{L}\right)^2 - 2\frac{n}{L}\right)(n-8) - \frac{n^3}{L^2}\right) + O(\lambda^4) \tag{D66}$$

$$= \left(\frac{n}{L}\right)^2 + \frac{\lambda^2}{8}\frac{n}{L}\left(1-\frac{n}{L}\right) + O(\lambda^4). \tag{D67}$$

Together, this gives

$$\left(\langle (H_l)^2\rangle - \langle H_lH_r\rangle\right) = \lambda^2\left(-\frac{1}{8}\frac{n}{L}\left(\frac{n}{L}-1\right) + \frac{1}{8}\frac{n}{L}\left(\frac{n}{L}-1\right)\right) + O(\lambda^4) = O(\lambda^4). \tag{D68}$$

For $L > 2$ we have

$$\langle H_lH_r\rangle = \langle 0|^{\otimes n}\left[\mathbb{1} + \frac{\lambda}{8}\sum_i X_i + \frac{\lambda^2}{16}\left(\frac{1}{3}\sum_{\langle i,j\rangle}X_iX_j + \frac{1}{8}\sum_{i\neq j\setminus\langle i,j\rangle}X_iX_j - \frac{1}{8}n\mathbb{1}\right)\right] \tag{D69}$$

$$\cdot\sum_{i=1}^{n_x/L}\sum_{j=1}^{n_y}\sum_{k=1}^{n_x/L}\sum_{l=1}^{n_y} Z_{Li,j}Z_{Li+1,j}Z_{Lk+1,l}Z_{Lk+2,l} \tag{D70}$$

$$\cdot\left[\mathbb{1} + \frac{\lambda}{8}\sum_i X_i + \frac{\lambda^2}{16}\left(\frac{1}{3}\sum_{\langle i,j\rangle}X_iX_j + \frac{1}{8}\sum_{i\neq j\setminus\langle i,j\rangle}X_iX_j - \frac{1}{8}n\mathbb{1}\right)\right]|0\rangle^{\otimes n} + \mathcal{O}(\lambda^3) \tag{D71}$$

$$= \left(\frac{n}{L}\right)^2 + \frac{\lambda^2}{64}\left(\left(\frac{n}{L}\right)(n-4) + \left(\left(\frac{n}{L}\right)^2 - \frac{n}{L}\right)(n-8) - \frac{n^3}{L^2}\right) + O(\lambda^4) \tag{D72}$$

$$= \left(\frac{n}{L}\right)^2 - \frac{\lambda^2}{8}\left(\left(\frac{n}{L}\right)^2 - \frac{1}{2}\frac{n}{L}\right) + O(\lambda^4). \tag{D73}$$

Together, this gives

$$\left(\langle (H_l)^2\rangle - \langle H_l H_r\rangle\right) = \frac{\lambda^2}{8}\left(-\frac{n}{L}\left(\frac{n}{L}-1\right) + \left(\frac{n}{L}\right)^2 - \frac{1}{2}\frac{n}{L}\right) + O(\lambda^4) = \frac{\lambda^2}{16}\frac{n}{L} + O(\lambda^4). \tag{D74}$$

Finally

$$\frac{\mathrm{Var}(H_2)}{\mathrm{Var}(\frac{1}{2}(H - H_l + H_r))} = 32L + O(\lambda^2). \tag{D75}$$

For $L > 2$ und $1 < i_r \leq L/2$,

$$\langle H_l H_r\rangle = \langle 0|^{\otimes n}\left[\mathbb{1} + \frac{\lambda}{8}\sum_i X_i + \frac{\lambda^2}{16}\left(\frac{1}{3}\sum_{\langle i,j\rangle} X_i X_j + \frac{1}{8}\sum_{i\neq j\setminus\langle i,j\rangle} X_i X_j - \frac{1}{8}n\mathbb{1}\right)\right] \tag{D76}$$

$$\cdot \sum_{i=1}^{n_x/L}\sum_{j=1}^{n_y}\sum_{k=1}^{n_x/L}\sum_{l=1}^{n_y} Z_{Li,j} Z_{Li+1,j} Z_{Lk+i_r,l} Z_{Lk+i_r+1,l} \tag{D77}$$

$$\cdot \left[\mathbb{1} + \frac{\lambda}{8}\sum_i X_i + \frac{\lambda^2}{16}\left(\frac{1}{3}\sum_{\langle i,j\rangle} X_i X_j + \frac{1}{8}\sum_{i\neq j\setminus\langle i,j\rangle} X_i X_j - \frac{1}{8}n\mathbb{1}\right)\right]|0\rangle^{\otimes n} + \mathcal{O}(\lambda^3) \tag{D78}$$

$$= \left(\frac{n}{L}\right)^2 + \frac{\lambda^2}{64}\left(\left(\frac{n}{L}\right)^2(n-8) - \frac{n^3}{L^2}\right) + O(\lambda^4) = \left(\frac{n}{L}\right)^2 - \frac{\lambda^2}{8}\left(\frac{n}{L}\right)^2 + O(\lambda^4). \tag{D79}$$

Together, this gives

$$\left(\langle (H_l)^2\rangle - \langle H_l H_r\rangle\right) = \frac{\lambda^2}{8}\left(-\frac{n}{L}\left(\frac{n}{L}-1\right) + \left(\frac{n}{L}\right)^2\right) + O(\lambda^4) = \frac{\lambda^2}{8}\frac{n}{L} + O(\lambda^4). \tag{D80}$$

Finally

$$\frac{\mathrm{Var}(H_2)}{\mathrm{Var}(\frac{1}{2}(H - H_l + H_r))} = 16L + O(\lambda^2). \tag{D81}$$

This means that it is beneficial to choose H_{cut} and H'_{cut} neighboring each other ($i_r = 1$). As this is possible in general, we only report the $i_r = 1$ results in Lemma 2 in the main text. For $L = 1$ we need to define

$$H_l = \sum_{i=1}^{n_x}\sum_{j=1}^{n_y} Z_{i,j} Z_{i+1,j} \qquad H_r = S H_l S = \sum_{i=1}^{n_x}\sum_{j=1}^{n_y} Z_{i,j} Z_{i,j+1}, \tag{D82}$$

$$H_l H_r = \sum_{i=1}^{n_x}\sum_{j=1}^{n_y}\sum_{k=1}^{n_x}\sum_{l=1}^{n_y} Z_{i,j} Z_{i+1,j} Z_{k,l} Z_{k+1,l}, \tag{D83}$$

$$\langle H_l H_r\rangle = \langle 0|^{\otimes n}\left[\mathbb{1} + \frac{\lambda}{8}\sum_i X_i + \frac{\lambda^2}{16}\left(\frac{1}{3}\sum_{\langle i,j\rangle} X_i X_j + \frac{1}{8}\sum_{i\neq j\setminus\langle i,j\rangle} X_i X_j - \frac{1}{8}n\mathbb{1}\right)\right] \tag{D84}$$

$$\cdot \sum_{i=1}^{n_x}\sum_{j=1}^{n_y}\sum_{k=1}^{n_x}\sum_{l=1}^{n_y} Z_{i,j} Z_{i+1,j} Z_{k,l} Z_{k+1,l} \tag{D85}$$

$$\cdot \left[\mathbb{1} + \frac{\lambda}{8} \sum_i X_i + \frac{\lambda^2}{16} \left(\frac{1}{3} \sum_{\langle i,j \rangle} X_i X_j + \frac{1}{8} \sum_{i \neq j \setminus \langle i,j \rangle} X_i X_j - \frac{1}{8} n \mathbb{1} \right) \right] |0\rangle^{\otimes n} + \mathcal{O}(\lambda^3) \tag{D86}$$

$$= n^2 + \frac{\lambda^2}{64} \left(2n(n-4) + ((n^2 - 2n)(n-8) - n^3 \right) + O(\lambda^4) \tag{D87}$$

$$= n^2 + \frac{\lambda^2}{8} n\,(1-n) + O(\lambda^4). \tag{D88}$$

Thus

$$\left(\langle (H_l)^2 \rangle - \langle H_l H_r \rangle \right) = \lambda^2 \left(-\frac{n}{8}\,(n-1) + \frac{n}{8}\,(n-1) \right) + O(\lambda^4) = O(\lambda^4) \tag{D89}$$

3. List of all used partitions

Here we list the exact partitions of our numerical results in Figs. 2, 4, 6, 7 and 8.

TABLE I. Explicit partitions of the transverse field XY model example, Fig. 2(a), and hard-core Bose-Hubbard model example, Fig. 4(a). To recall, all examples apart from the spinful Hubbard model are on a $(n_x, n_y) = (4, 6)$ rectangular lattice with periodic boundary conditions; thus all indices are modulo n_x and respectively n_y. For the geometric partitionings we show in Fig. 2(a) the sampling improvements $\mathcal{G}_{|E_0\rangle}(\mathcal{B}_{\text{Pauli}}, \mathcal{B}_{L_x,L_y})$, and for Pauli partitioning the combined variance of the partition for Pauli measurements $[\sum_{H_b \in \mathcal{B}_{\text{Pauli}}} \sqrt{\text{Var}_{|\psi\rangle}(H_b)}]^2$ is given. For the transverse field XY model, Eq. (21), $J_X = \frac{1}{2}(1+\eta), J_Y = \frac{1}{2}(1-\eta)$, and $h_Z = h$. Similarly, for the Bose-Hubbard model, when allowing maximally one boson per lattice side, Eq. (26), one can map it to a spin basis Eq. (25), where it has transverse field XY model form with $J_X = J_Y = \frac{1}{2}J$, and $h_Z = -\frac{1}{2}h$.

Label	Color	Partitioning
Pauli		$H_1 = -\sum_{i=1}^{n_x=4}\sum_{j=1}^{n_x=6} J_X \left(X_{i,j}X_{i+1,j} + X_{i,j}X_{i,j+1}\right) \quad H_3 = -\sum_{i=1}^{n_x=4}\sum_{j=1}^{n_x=6} h_Z Z_{i,j}$ $H_2 = -\sum_{i=1}^{n_x=4}\sum_{j=1}^{n_x=6} J_Y \left(Y_{i,j}Y_{i+1,j} + Y_{i,j}Y_{i,j+1}\right)$ (D90)
$L = 1$, $(L_x, L_y) = (n_x, 1)$		$H_1 = -\sum_{i=1}^{n_x=4}\sum_{j=1}^{n_x=6} \left(J_X X_{i,j}X_{i+1,j} + J_Y Y_{i,j}Y_{i+1,j}\right) - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_x=6} \frac{h_Z}{2} Z_{i,j}$ $H_2 = -\sum_{i=1}^{n_x=4}\sum_{j=1}^{n_x=6} \left(J_X X_{i,j}X_{i,j+1} + J_Y Y_{i,j}Y_{i,j+1}\right) - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_x=6} \frac{h_Z}{2} Z_{i,j}$ (D91)
$L = 2$, $(L_x, L_y) = (n_x, 2)$		$H_1 = -\sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y/L=3} \left(J_X X_{i,2j}X_{i,2j+1} + J_Y Y_{i,2j}Y_{i,2j+1}\right)$ $-\sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=6} \left(\frac{J_X}{2}X_{i,j}X_{i+1,j} + \frac{J_Y}{2}Y_{i,j}Y_{i+1,j}\right) - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=6} \frac{h_Z}{2} Z_{i,j}$ $H_2 = -\sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y/L=3} \left(J_X X_{i,2j-1}X_{i,2j} - J_Y Y_{i,2j-1}Y_{i,2j}\right)$ $-\sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=6} \left(\frac{J_X}{2}X_{i,j}X_{i+1,j} + \frac{J_Y}{2}Y_{i,j}Y_{i+1,j}\right) - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=6} \frac{h_Z}{2} Z_{i,j}$ (D92)

TABLE I. (Continued.)

Label	Color	Partitioning
$L = 3$, $(L_x, L_y) = (n_x, 3)$		$H_1 = -\sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y/L=2}\left[J_X\left(X_{i,3j}X_{i,3j+1} + \frac{1}{2}X_{i,3j+1}X_{i,3j+2}\right) + J_Y\left(Y_{i,3j}Y_{i,3j+1} + \frac{1}{2}Y_{i,3j+1}Y_{i,3j+2}\right)\right]$ $-\sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=6}\left(\frac{J_X}{2}X_{i,j}X_{i+1,j}\frac{J_Y}{2}Y_{i,j}Y_{i+1,j}\right) - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=6}\frac{h_Z}{2}Z_{i,j}$ $H_2 = -\sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y/L=2}\left[J_X\left(X_{i,3j-1}X_{i,3j} + \frac{1}{2}X_{i,3j+1}X_{i,3j+2}\right) + J_Y\left(Y_{i,3j-1}Y_{i,3j} + \frac{1}{2}Y_{i,3j+1}Y_{i,3j+2}\right)\right]$ $-\sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=6}\left(\frac{J_X}{2}X_{i,j}X_{i+1,j}\frac{J_Y}{2}Y_{i,j}Y_{i+1,j}\right) - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=6}\frac{h_Z}{2}Z_{i,j}$ (D93)
$(L_x, L_y) =$ 2-local		$H_1 = -\sum_{i=1}^{n_x/2=2}\sum_{j=1}^{n_y=6}\left(J_X X_{2i,j}X_{2i+1,j} + J_Y Y_{2i,j}Y_{2i+1,j}\right) - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=6}\frac{h_Z}{4}Z_{i,j}$ $H_2 = -\sum_{i=1}^{n_x/2=2}\sum_{j=1}^{n_y=6}\left(J_X X_{2i-1,j}X_{2i,j} + J_Y Y_{2i-1,j}Y_{2i,j}\right) - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=6}\frac{h}{4}X_{i,j}$ $H_3 = -\sum_{i=1}^{n_x/2=2}\sum_{j=1}^{n_y=6}\left(J_X X_{i,2j}X_{i,2j+1} + J_Y Y_{i,2j}Y_{i,2j+1}\right) - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=6}\frac{h_Z}{4}Z_{i,j}$ $H_4 = -\sum_{i=1}^{n_x/2=2}\sum_{j=1}^{n_y=6}\left(J_X X_{2j-1}X_{i,2j} + J_Y Y_{i,2j-1}Y_{i,2j}\right) - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=6}\frac{h}{4}X_{i,j}$ (D94)
$(L_x, L_y) = (2, 2)$		$H_1 = -\sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y/L_y=3}\left(J_X X_{i,2j}X_{i,2j+1} + (J_Y Y_{i,2j}Y_{i,2j+1}\right)$ $-\sum_{i=1}^{n_x/L_x=2}\sum_{j=1}^{n_y=6}\left(J_X X_{2i,j}X_{2i+1,j} + J_Y Y_{2i,j}Y_{2i+1,j}\right) - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=6}\frac{h_Z}{2}Z_{i,j}$ $H_2 = -\sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y/L_y=3}\left(J_X X_{i,2j-1}X_{i,2j} + (J_Y Y_{i,2j-1}Y_{i,2j}\right)$ $-\sum_{i=1}^{n_x/L_x=2}\sum_{j=1}^{n_y=6}\left(J_X X_{2i-1,j}X_{2i,j} + J_Y Y_{2i-1,j}Y_{2i,j}\right) - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=6}\frac{h_Z}{2}Z_{i,j}$ (D95)
$(L_x, L_y) = (2, 3)$		$H_1 = -\sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y/L=2}\left[J_X\left(X_{i,3j}X_{i,3j+1} + \frac{1}{2}X_{i,3j+1}X_{i,3j+2}\right) + J_Y\left(Y_{i,3j}Y_{i,3j+1} + \frac{1}{2}Y_{i,3j+1}Y_{i,3j+2}\right)\right]$ $-\sum_{i=1}^{n_x/L_x=2}\sum_{j=1}^{n_y=6}\left(\frac{J_X}{2}X_{2i,j}X_{2i+1,j}\frac{J_Y}{2}Y_{2i,j}Y_{2i+1,j}\right) - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=6}\frac{h_Z}{2}Z_{i,j}$ $H_2 = -\sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y/L=2}\left[J_X\left(X_{i,3j-1}X_{i,3j} + \frac{1}{2}X_{i,3j+1}X_{i,3j+2}\right) + J_Y\left(Y_{i,3j-1}Y_{i,3j} + \frac{1}{2}Y_{i,3j+1}Y_{i,3j+2}\right)\right]$ $-\sum_{i=1}^{n_x/L_x=2}\sum_{j=1}^{n_y=6}\left(\frac{J_X}{2}X_{2i-1,j}X_{2i,j}\frac{J_Y}{2}Y_{2i-1,j}Y_{2i,j}\right) - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=6}\frac{h_Z}{2}Z_{i,j}$ (D96)

TABLE II. Explicit partitions of the transverse field Ising model example, Fig. 2(b). To recall, all examples apart from the spinful Hubbard model are on a $(n_x, n_y) = (4, 6)$ rectangular lattice with periodic boundary conditions; thus all indices are modulo n_x and respectively n_y. For the geometric partitionings we show in Fig. 2(b) the sampling improvements $\mathcal{G}_{|E_0\rangle}(\mathcal{B}_{\text{Pauli}}, \mathcal{B}_{L_x,L_y})$, and for Pauli partitioning the combined variance of the partition for Pauli measurements $[\sum_{H_b \in \mathcal{B}_{\text{Pauli}}} \sqrt{\text{Var}_{|\psi\rangle}(H_b)}]^2$ is given.

Label	Color	Partitioning
Pauli		$H_1 = -\sum_{i=1}^{n_x=4}\sum_{j=1}^{n_x=6} J\left(Z_{i,j}Z_{i+1,j} + Z_{i,j}Z_{i,j+1}\right) \quad H_2 = -\sum_{i=1}^{n_x=4}\sum_{j=1}^{n_x=6} hX_{i,j}$ (D97)
$L = 1$, $(L_x, L_y) = (n_x, 1)$		$H_1 = -\sum_{i=1}^{n_x=4}\sum_{j=1}^{n_x=6} JZ_{i,j}Z_{i+1,j} - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_x=6}\frac{h}{2}X_{i,j} \quad H_2 = -\sum_{i=1}^{n_x=4}\sum_{j=1}^{n_x=6} JZ_{i,j}Z_{i,j+1} - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_x=6}\frac{h}{2}X_{i,j}$ (D98)
$L = 2$, $(L_x, L_y) = (n_x, 2)$		$H_1 = -\sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=6}\frac{J}{2}Z_{i,j}Z_{i+1,j} - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y/L=3} JZ_{i,2j}Z_{i,2j+1} - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=6}\frac{h}{2}X_{i,j}$ $H_2 = -\sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=6}\frac{J}{2}Z_{i,j}Z_{i+1,j} - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y/L=3} JZ_{i,2j-1}Z_{i,2j} - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=6}\frac{h}{2}X_{i,j}$ (D99)
$L = 3$, $(L_x, L_y) = (n_x, 3)$		$H_1 = -\sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=6}\frac{J}{2}Z_{i,j}Z_{i+1,j} - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y/L=2} J\left(Z_{i,3j}Z_{i,3j+1} + \frac{1}{2}Z_{i,3j+1}Z_{i,3j+2}\right) - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=6}\frac{h}{2}X_{i,j}$ $H_2 = -\sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=6}\frac{J}{2}Z_{i,j}Z_{i+1,j} - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y/L=2} J\left(Z_{i,3j-1}Z_{i,3j} + \frac{1}{2}Z_{i,3j+1}Z_{i,3j+2}\right) - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=6}\frac{h}{2}X_{i,j}$ (D100)
$(L_x, L_y) = $ 2-local		$H_1 = -\sum_{i=1}^{n_x/2=2}\sum_{j=1}^{n_y=6} JZ_{2i,j}Z_{2i+1,j} - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=6}\frac{h}{4}X_{i,j} \quad H_2 = -\sum_{i=1}^{n_x/2=2}\sum_{j=1}^{n_y=6} JZ_{2i-1,j}Z_{2i,j} - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=6}\frac{h}{4}X_{i,j}$ $H_3 = -\sum_{i=1}^{n_x/2=2}\sum_{j=1}^{n_y=6} JZ_{i,2j}Z_{i,2j+1} - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=6}\frac{h}{4}X_{i,j} \quad H_4 = -\sum_{i=1}^{n_x/2=2}\sum_{j=1}^{n_y=6} JZ_{i,2j-1}Z_{i,2j} - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=6}\frac{h}{4}X_{i,j}$ (D101)
$(L_x, L_y) = (2, 2)$		$H_1 = -\sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y/L_y=3} JZ_{i,2j}Z_{i,2j+1} - \sum_{i=1}^{n_x/L_x=2}\sum_{j=1}^{n_y=6} JZ_{2i,j}Z_{2i+1,j} - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=6}\frac{h}{2}X_{i,j}$ $H_2 = -\sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y/L_y=3} JZ_{i,2j-1}Z_{i,2j} - \sum_{i=1}^{n_x/L_x=2}\sum_{j=1}^{n_y=6} JZ_{2i-1,j}Z_{2i,j} - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=6}\frac{h}{2}X_{i,j}$ (D102)
$(L_x, L_y) = (2, 3)$		$H_1 = -\sum_{i=1}^{n_x/L_x=2}\sum_{j=1}^{n_y=6} JZ_{2i,j}Z_{2i+1,j} - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y/L=2} J\left(Z_{i,3j}Z_{i,3j+1} + \frac{1}{2}Z_{i,3j+1}Z_{i,3j+2}\right) - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=6}\frac{h}{2}X_{i,j}$ $H_2 = -\sum_{i=1}^{n_x/L_x=2}\sum_{j=1}^{n_y=6} JZ_{2i-1,j}Z_{2i,j} - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y/L=2} J\left(Z_{i,3j-1}Z_{i,3j} + \frac{1}{2}Z_{i,3j+1}Z_{i,3j+2}\right) - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=6}\frac{h}{2}X_{i,j}$ (D103)

TABLE III. Explicit partitions of the transverse field biaxial next-nearest-neighbor Ising example, Fig. 2(c). To recall, all examples apart from the spinful Hubbard model are on a $(n_x, n_y) = (4, 6)$ rectangular lattice with periodic boundary conditions; thus all indices are modulo n_x and respectively n_y. For the geometric partitionings we show in Fig. 2(c) the sampling improvements $\mathcal{G}_{|E_0\rangle}(\mathcal{B}_{\text{Pauli}}, \mathcal{B}_{L_x,L_y})$, and for Pauli partitioning the combined variance of the partition for Pauli measurements $[\sum_{H_b \in \mathcal{B}_{\text{Pauli}}} \sqrt{\text{Var}_{|\psi\rangle}(H_b)}]^2$ is given. Further, for the transverse field biaxial next-nearest-neighbor Ising model the unit cell will be axial 3-local, and hence partitions $L = 2$, $(L_x, L_y) = (2, 2)$, and $(L_x, L_y) = (2, 3)$ are not applicable. Similarly, a bipartition in $H = H_1 + H_2$ will again be possible for $L_x, L_y \geq 5$, where then the unit cell is shifted by 2 due to the next-nearest-neighbor coupling.

Label	Color	Partitioning
Pauli		$H_1 = -\sum_{i=1}^{n_x=4}\sum_{j=1}^{n_x=6} J\left[\left(Z_{i,j}Z_{i+1,j} + Z_{i,j}Z_{i,j+1}\right) - \kappa J\left(Z_{i,j}Z_{i+2,j} + Z_{i,j}Z_{i,j+2}\right)\right]$ $H_2 = -\sum_{i=1}^{n_x=4}\sum_{j=1}^{n_x=6} hX_{i,j}$ (D104)
$L = 1$, $(L_x, L_y) = (n_x, 1)$		$H_1 = -\sum_{i=1}^{n_x=4}\sum_{j=1}^{n_x=6} J\left(Z_{i,j}Z_{i+1,j} - \kappa Z_{i,j}Z_{i+2,j}\right) - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_x=6}\frac{h}{2}X_{i,j}$ $H_2 = -\sum_{i=1}^{n_x=4}\sum_{j=1}^{n_x=6} J\left(Z_{i,j}Z_{i,j+1} - \kappa Z_{i,j}Z_{i,j+2}\right) - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_x=6}\frac{h}{2}X_{i,j}$ (D105)
$L = 3$, $(L_x, L_y) = (n_x, 3)$		$H_1 = -\sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y/L=2} J\left(\frac{1}{2}Z_{i,3j}Z_{i,3j+1} + \frac{1}{2}Z_{i,3j+1}Z_{i,3j+2} - \kappa Z_{i,3j}Z_{i,3j+2}\right)$ $- \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=6}\frac{J}{3}\left(Z_{i,j}Z_{i+1,j} - \kappa Z_{i,j}Z_{i+2,j}\right) - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=6}\frac{h}{3}X_{i,j}$ $H_2 - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y/L=2} J\left(\frac{1}{2}Z_{i,3j+1}Z_{i,3j+2} + \frac{1}{2}Z_{i,3j+2}Z_{i,3j+3} - \kappa Z_{i,3j+1}Z_{i,3j+3}\right)$ $- \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=6}\frac{J}{3}\left(Z_{i,j}Z_{i+1,j} - \kappa Z_{i,j}Z_{i+2,j}\right) - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=6}\frac{h}{3}X_{i,j}$ $H_3 - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y/L=2} J\left(\frac{1}{2}Z_{i,3j-1}Z_{i,3j} + \frac{1}{2}Z_{i,3j}Z_{i,3j+1} - \kappa Z_{i,3j-1}Z_{i,3j+1}\right)$ $- \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=6}\frac{J}{3}\left(Z_{i,j}Z_{i+1,j} - \kappa Z_{i,j}Z_{i+2,j}\right) - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=6}\frac{h}{3}X_{i,j}$ (D106)
$(L_x, L_y) = 3$-local		$H_1 = -\sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y/3=2} J\left(\frac{1}{2}Z_{i,3j}Z_{i,3j+1} + \frac{1}{2}Z_{i,3j+1}Z_{i,3j+2} - \kappa Z_{i,3j}Z_{i,3j+2}\right) - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=6}\frac{h}{6}X_{i,j}$ $H_2 = -\sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y/3=2} J\left(\frac{1}{2}Z_{i,3j+1}Z_{i,3j+2} + \frac{1}{2}Z_{i,3j+2}Z_{i,3j+3} - \kappa Z_{i,3j+1}Z_{i,3j+3}\right) - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=6}\frac{h}{6}X_{i,j}$ $H_{2+k} = -\sum_{j=1}^{n_y/3=2} J\left(\frac{1}{2}Z_{k,3j-1}Z_{1,3j} + \frac{1}{2}Z_{k,3j}Z_{k,3j+1} - \kappa Z_{k,3j-1}Z_{k,3j+1}\right)$ $- \sum_{j=1}^{n_y=6} J\left(\frac{1}{2}Z_{k+1,j}Z_{k+2,j} + \frac{1}{2}Z_{k+2,j}Z_{k+3,j} - \kappa Z_{k+1,j}Z_{k+3,j}\right) - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=6}\frac{h}{6}X_{i,j}$ $H_3 = H_{2+k=1} \quad H_4 = H_{2+k=2} \quad H_5 = H_{2+k=3} \quad H_6 = H_{2+k=4}$ (D107)

TABLE IV. Explicit partitions of the spinless Hubbard example, Fig. 4(b). To recall, all examples apart from the spinful Hubbard model are on a $(n_x, n_y) = (4, 6)$ rectangular lattice with periodic boundary conditions; thus all indices are modulo n_x and respectively n_y. For the geometric partitionings we show in Fig. 4(b) the sampling improvements $\mathcal{G}_{|E_0\rangle}(\mathcal{B}_{\text{Pauli}}, \mathcal{B}_{L_x,L_y})$, and for Pauli partitioning the combined variance of the partition for Pauli measurements $[\sum_{H_b\in\mathcal{B}_{\text{Pauli}}} \sqrt{\text{Var}_{|\psi\rangle}(H_b)}]^2$ is given.

Label	Color	Partitioning
Pauli		$H_1 = -\sum_{i=1}^{n_x/2=2}\sum_{j=1}^{n_y/2=3} t\left(c_{i,j}^\dagger c_{i+1,j} + c_{i+1,j}^\dagger c_{i,j}\right)$ $\quad H_2 = -\sum_{i=1}^{n_x/2=2}\sum_{j=1}^{n_y/2=3} t\left(c_{i,j}^\dagger c_{i+1,j} + c_{i+1,j}^\dagger c_{i,j}\right)$ $H_3 = -\sum_{i=1}^{n_x/2=2}\sum_{j=1}^{n_y/2=3} t\left(c_{i,j}^\dagger c_{i+1,j} + c_{i+1,j}^\dagger c_{i,j}\right)$ $\quad H_4 = -\sum_{i=1}^{n_x/2=2}\sum_{j=1}^{n_y/2=3} t\left(c_{i,j}^\dagger c_{i+1,j} + c_{i+1,j}^\dagger c_{i,j}\right)$ $H_5 = +\sum_{i=1}^{n_x=4}\sum_{j=1}^{n_x=6} U(n_{i,j}n_{i,j+1} + n_{i,j}n_{i+1,j}) - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=6}\mu n_{i,j}$ (D108)
$L = 1, (L_x, L_y) = (n_x, 1)$		$H_1 = -\sum_{i=1}^{n_x=4}\sum_{j=1}^{n_x=6}\left(tc_{i,j}^\dagger c_{i+1,j} + tc_{i+1,j}^\dagger c_{i,j} - Un_{i,j}n_{i+1,j}\right) - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=6}\frac{\mu}{2}n_{i,j}$ $H_2 = -\sum_{i=1}^{n_x=4}\sum_{j=1}^{n_x=6}\left(tc_{i,j}^\dagger c_{i,j+1} + tc_{i,j+1}^\dagger c_{i,j} - Un_{i,j}n_{i,j+1}\right) - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=6}\frac{\mu}{2}n_{i,j}$ (D109)
$L = 2, (L_x, L_y) = (n_x, 2)$		$H_1 = -\sum_{i=1}^{n_x=4}\sum_{j=1}^{n_x=6}\left(\frac{t}{2}c_{i,j}^\dagger c_{i+1,j} + \frac{t}{2}c_{i+1,j}^\dagger c_{i,j} - \frac{U}{2}n_{i,j}n_{i+1,j}\right)$ $-\sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y/L=3}\left(tc_{i,2j}^\dagger c_{i,2j+1} + tc_{i,2j+1}^\dagger c_{i,2j} - Un_{i,2j}n_{i,2j+1}\right) - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=6}\frac{\mu}{2}n_{i,j}$ $H_2 = -\sum_{i=1}^{n_x=4}\sum_{j=1}^{n_x=6}\left(\frac{t}{2}c_{i,j}^\dagger c_{i+1,j} + \frac{t}{2}c_{i+1,j}^\dagger c_{i,j} - \frac{U}{2}n_{i,j}n_{i+1,j}\right)$ $-\sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y/L=3}\left(tc_{i,2j-1}^\dagger c_{i,2j} + tc_{i,2j}^\dagger c_{i,2j-1} - Un_{i,2j-1}n_{i,2j}\right) - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=6}\frac{\mu}{2}n_{i,j}$ (D110)
$L = 3, (L_x, L_y) = (n_x, 3)$		$H_1 = -\sum_{i=1}^{n_x=4}\sum_{j=1}^{n_x=6}\left(\frac{t}{2}c_{i,j}^\dagger c_{i+1,j} + \frac{t}{2}c_{i+1,j}^\dagger c_{i,j} - \frac{U}{2}n_{i,j}n_{i+1,j}\right)$ $-\sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y/L=2}\left(tc_{i,3j}^\dagger c_{i,3j+1} + tc_{i,3j+1}^\dagger c_{i,3j} - Un_{i,3j}n_{i,3j+1}\right)$ $-\sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y/L=2}\frac{1}{2}\left(tc_{i,3j+1}^\dagger c_{i,3j+2} + tc_{i,3j+2}^\dagger c_{i,3j+1} - Un_{i,3j+1}n_{i,3j+2}\right) - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=6}\frac{\mu}{2}n_{i,j}$ $H_2 - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_x=6}\left(\frac{t}{2}c_{i,j}^\dagger c_{i+1,j} + \frac{t}{2}c_{i+1,j}^\dagger c_{i,j} - \frac{U}{2}n_{i,j}n_{i+1,j}\right)$ $-\sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y/L=2}\left(tc_{i,3j-1}^\dagger c_{i,3j} + tc_{i,3j}^\dagger c_{i,3j-1} - Un_{i,3j-1}n_{i,3j}\right)$ $-\sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y/L=2}\frac{1}{2}\left(tc_{i,3j-1}^\dagger c_{i,3j-2} + tc_{i,3j-2}^\dagger c_{i,3j-1} - Un_{i,3j-1}n_{i,3j-2}\right) - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=6}\frac{\mu}{2}n_{i,j}$ (D111)

TABLE IV. Continued.

Label	Color	Partitioning
$(L_x, L_y) = 2$-local		$H_1 = -\sum_{i=1}^{n_x/2=2}\sum_{j=1}^{n_y=6}\left(tc^\dagger_{2i,j}c_{2i+1,j} + tc^\dagger_{2i+1,j}c_{2i,j} - Un_{2i,j}n_{2i+1,j}\right) - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=6}\frac{\mu}{4}n_{i,j}$ $H_2 = -\sum_{i=1}^{n_x/2=2}\sum_{j=1}^{n_y=6}\left(tc^\dagger_{2i-1,j}c_{2i,j} + tc^\dagger_{2i,j}c_{2i-1,j} - Un_{2i-1,j}n_{2i,j}\right) - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=6}\frac{\mu}{4}n_{i,j}$ $H_3 = -\sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y/2=3}\left(tc^\dagger_{i,2j}c_{i,2j+1} + tc^\dagger_{i,2j+1}c_{i,2j} - Un_{i,2j}n_{i,2j+1}\right) - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=6}\frac{\mu}{4}n_{i,j}$ $H_4 = -\sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y/2=3}\left(tc^\dagger_{i,2j-1}c_{i,2j} + tc^\dagger_{i,2j}c_{i,2j-1} - Un_{i,2j-1}n_{i,2j}\right) - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=6}\frac{\mu}{4}n_{i,j}$ (D112)
$(L_x, L_y) = (2, 2)$		$H_1 = -\sum_{i=1}^{n_x/L_x=2}\sum_{j=1}^{n_x=6}\left(tc^\dagger_{2i,j}c_{2i+1,j} + tc^\dagger_{2i+1,j}c_{2i,j} - Un_{2i,j}n_{2i+1,j}\right)$ $-\sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y/L_y=3}\left(tc^\dagger_{i,2j}c_{i,2j+1} + tc^\dagger_{i,2j+1}c_{i,2j} - Un_{i,2j}n_{i,2j+1}\right) - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=6}\frac{\mu}{2}n_{i,j}$ $H_2 = -\sum_{i=1}^{n_x/L_x=2}\sum_{j=1}^{n_x=6}\left(tc^\dagger_{2i-1,j}c_{2i,j} + tc^\dagger_{2i,j}c_{2i-1,j} - Un_{2i-1,j}n_{2i,j}\right)$ $-\sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y/L=3}\left(tc^\dagger_{i,2j-1}c_{i,2j} + tc^\dagger_{i,2j}c_{i,2j-1} - Un_{i,2j-1}n_{i,2j}\right) - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=6}\frac{\mu}{2}n_{i,j}$ (D113)
$(L_x, L_y) = (2, 3)$		$H_1 = -\sum_{i=1}^{n_x/L_x=2}\sum_{j=1}^{n_x=6}\left(tc^\dagger_{2i,j}c_{2i+1,j} + tc^\dagger_{2i+1,j}c_{2i,j} - Un_{2i,j}n_{2i+1,j}\right)$ $-\sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y/L_y=2}\left(tc^\dagger_{i,3j}c_{i,3j+1} + tc^\dagger_{i,3j+1}c_{i,3j} - Un_{i,3j}n_{i,3j+1}\right)$ $-\sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y/L_y=2}\frac{1}{2}\left(tc^\dagger_{i,3j+1}c_{i,3j+2} + tc^\dagger_{i,3j+2}c_{i,3j+1} - Un_{i,3j+1}n_{i,3j+2}\right) - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=6}\frac{\mu}{2}n_{i,j}$ $H_2 = -\sum_{i=1}^{n_x/L_x=2}\sum_{j=1}^{n_x=6}\left(tc^\dagger_{2i-1,j}c_{2i,j} + tc^\dagger_{2i,j}c_{2i-1,j} - Un_{2i-1,j}n_{2i,j}\right)$ $-\sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y/L=2}\left(tc^\dagger_{i,3j-1}c_{i,3j} + tc^\dagger_{i,3j}c_{i,3j-1} - Un_{i,3j-1}n_{i,3j}\right)$ $-\sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y/L=2}\frac{1}{2}\left(tc^\dagger_{i,3j-1}c_{i,3j-2} + tc^\dagger_{i,3j-2}c_{i,3j-1} - Un_{i,3j-1}n_{i,3j-2}\right) - \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=6}\frac{\mu}{2}n_{i,j}$ (D114)

TABLE V. Explicit partitions of the spinful Hubbard example, Figs. 4(c) and 7. The spinful Hubbard model example uses a bi-layer $(n_x, n_y) = (4, 3, 2)$ rectangular lattice with periodic boundary conditions in n_x and n_y direction; thus all indices in the first two indices are modulo n_x and respectively n_y. For the geometric partitionings we show in Fig. 4(c) the sampling improvements $\mathcal{G}_{|E_0\rangle}(\mathcal{B}_{\text{Pauli}}, \mathcal{B}_{L_x,L_y})$, and for Pauli partitioning the combined variance of the partition for Pauli measurements $[\sum_{H_b\in\mathcal{B}_{\text{Pauli}}} \sqrt{\text{Var}_{|\psi\rangle}(H_b)}]^2$ is given.

Label	Color	Partitioning
Pauli		$H_1 = -\sum_{i=1}^{n_x/2=2} \sum_{\sigma\in\{\uparrow,\downarrow\}} t\left(c^\dagger_{2i,1,\sigma} c_{2i+1,1,\sigma} + c^\dagger_{2i+1,1} c_{2i,1}\right) - \sum_{i=1}^{n_x=4} \sum_{\sigma\in\{\uparrow,\downarrow\}} t\left(c^\dagger_{i,2,\sigma} c_{i,3,\sigma} + c^\dagger_{i,3} c_{i,2}\right)$ $H_2 = -\sum_{i=1}^{n_x/2=2} \sum_{\sigma\in\{\uparrow,\downarrow\}} t\left(c^\dagger_{2i,2,\sigma} c_{2i+1,2,\sigma} + c^\dagger_{2i+1,2} c_{2i,2}\right) - \sum_{i=1}^{n_x=4} \sum_{\sigma\in\{\uparrow,\downarrow\}} t\left(c^\dagger_{i,3,\sigma} c_{i,1,\sigma} + c^\dagger_{i,1} c_{i,3}\right)$ $H_3 = -\sum_{i=1}^{n_x/2=2} \sum_{\sigma\in\{\uparrow,\downarrow\}} t\left(c^\dagger_{2i,3,\sigma} c_{2i+1,3,\sigma} + c^\dagger_{2i+1,3} c_{2i,3}\right) - \sum_{i=1}^{n_x=4} \sum_{\sigma\in\{\uparrow,\downarrow\}} t\left(c^\dagger_{i,1,\sigma} c_{i,2,\sigma} + c^\dagger_{i,2} c_{i,1}\right)$ $H_4 = -\sum_{i=1}^{n_x/2=2} \sum_{j=1}^{n_y=3} \sum_{\sigma\in\{\uparrow,\downarrow\}} t\left(c^\dagger_{2i-1,j} c_{2i,j} + c^\dagger_{2i,j} c_{2i-1,j}\right)$ $H_5 = +\sum_{i=1}^{n_x=4} \sum_{j=1}^{n_x=3} U n_{i,j,\uparrow} n_{i,j,\downarrow} - \mu(n_{i,j,\uparrow} + n_{i,j,\downarrow})$ (D115)
$(L_x, L_y, L_{\uparrow,\downarrow}) =$ $(n_x, n_y, 1)$, & $(1, 1, n_{\uparrow,\downarrow})$		$H_1 = -\sum_{\sigma\in\{\uparrow,\downarrow\}} \left[\sum_{i=1}^{n_x=4} \sum_{j=1}^{n_y=3} \left(t c^\dagger_{i,j} c_{i+1,j} + t c^\dagger_{i+1,j} c_{i,j} + t c^\dagger_{i,j} c_{i,j+1} + t c^\dagger_{i,j+1} c_{i,j}\right) + \frac{\mu}{2} n_{ij,\sigma}\right]$ $H_2 = +\sum_{i=1}^{n_x=4} \sum_{j=1}^{n_x=3} U n_{i,j,\uparrow} n_{i,j,\downarrow} - \frac{\mu}{2}(n_{i,j,\uparrow} + n_{i,j,\downarrow})$ (D116)
$(L_x, L_y, L_{\uparrow,\downarrow}) =$ $(2, n_y, n_{\uparrow,\downarrow})$		$H_1 = \sum_{i=1}^{n_x=4} \sum_{j=1}^{n_y=3} \left[\sum_{\sigma\in\{\uparrow,\downarrow\}} -\left(\frac{t}{2} c^\dagger_{i,j,\sigma} c_{i,j+1,\sigma} + \frac{t}{2} c^\dagger_{i,j+1,\sigma} c_{i,j,\sigma}\right) - \frac{\mu}{2} n_{i,j,\sigma}\right] + \frac{U}{2} n_{i,j,\uparrow} n_{i,j,\downarrow}$ $- \sum_{i=1}^{n_x/L_x=2} \sum_{j=1}^{n_y=3} \sum_{\sigma\in\{\uparrow,\downarrow\}} t c^\dagger_{2i,j,\sigma} c_{2i+1,j,\sigma} + t c^\dagger_{2i+1,j,\sigma} c_{2i,j,\sigma}$ $H_2 = \sum_{i=1}^{n_x=4} \sum_{j=1}^{n_y=3} \left[\sum_{\sigma\in\{\uparrow,\downarrow\}} -\left(\frac{t}{2} c^\dagger_{i,j,\sigma} c_{i,j+1,\sigma} + \frac{t}{2} c^\dagger_{i,j+1,\sigma} c_{i,j,\sigma}\right) - \frac{\mu}{2} n_{i,j,\sigma}\right] + \frac{U}{2} n_{i,j,\uparrow} n_{i,j,\downarrow}$ $- \sum_{i=1}^{n_x/L_x=2} \sum_{j=1}^{n_y=3} \sum_{\sigma\in\{\uparrow,\downarrow\}} t c^\dagger_{2i-1,j,\sigma} c_{2i,j,\sigma} + t c^\dagger_{2i,j,\sigma} c_{2i-1,j,\sigma}$ (D117)
$(L_x, L_y, L_{\uparrow,\downarrow}) =$ $(n_x, 1, 1)$, $(1, n_y, 1)$, & $(1, 1, n_{\uparrow,\downarrow})$		$H_1 = -\sum_{i=1}^{n_x=4} \sum_{j=1}^{n_y=3} \sum_{\sigma\in\{\uparrow,\downarrow\}} \left(t c^\dagger_{i,j,\sigma} c_{i+1,j,\sigma} + t c^\dagger_{i+1,j,\sigma} c_{i,j,\sigma}\right) - \frac{\mu}{3} n_{i,j,\sigma}$ $H_2 = -\sum_{i=1}^{n_x=4} \sum_{j=1}^{n_y=3} \sum_{\sigma\in\{\uparrow,\downarrow\}} \left(t c^\dagger_{i,j,\sigma} c_{i,j+1,\sigma} + t c^\dagger_{i,j+1,\sigma} c_{i,j,\sigma}\right) - \frac{\mu}{3} n_{i,j,\sigma}$ $H_3 = +\sum_{i=1}^{n_x=4} \sum_{j=1}^{n_y=3} U n_{i,j,\uparrow} n_{i,j,\downarrow} - \frac{\mu}{3}\left(n_{i,j,\uparrow} + n_{i,j,\downarrow}\right)$ (D118)

TABLE V. Continued.

Label	Color	Partitioning
$(L_x, L_y, L_{\uparrow,\downarrow}) = (1, n_y, n_{\uparrow,\downarrow})$, & $(n_x, 1, 1)$		$H_1 = \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=3}\left[\sum_{\sigma\in\{\uparrow,\downarrow\}} -\left(tc^\dagger_{i,j,\sigma}c_{i,j+1,\sigma} + tc^\dagger_{i,j+1,\sigma}c_{i,j,\sigma}\right) - \frac{\mu}{2}n_{i,j,\sigma}\right] + Un_{i,j,\uparrow}n_{i,j,\downarrow}$ $H_2 = \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=3}\sum_{\sigma\in\{\uparrow,\downarrow\}} -\left(tc^\dagger_{i,j,\sigma}c_{i+1,j,\sigma} + tc^\dagger_{i+1,j,\sigma}c_{i,j,\sigma}\right) - \frac{\mu}{2}n_{i,j,\sigma}$ (D119)
$(L_x, L_y, L_{\uparrow,\downarrow}) = (n_x, 1, n_{\uparrow,\downarrow})$, & $(1, n_y, 1)$		$H_1 = \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=3}\left[\sum_{\sigma\in\{\uparrow,\downarrow\}} -\left(tc^\dagger_{i,j,\sigma}c_{i+1,j,\sigma} + tc^\dagger_{i+1,j,\sigma}c_{i,j,\sigma}\right) - \frac{\mu}{2}n_{i,j,\sigma}\right] + Un_{i,j,\uparrow}n_{i,j,\downarrow}$ $H_2 = \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=3}\sum_{\sigma\in\{\uparrow,\downarrow\}} -\left(tc^\dagger_{i,j,\sigma}c_{i,j+1,\sigma} + tc^\dagger_{i,j+1,\sigma}c_{i,j,\sigma}\right) - \frac{\mu}{2}n_{i,j,\sigma}$ (D120)
$(L_x, L_y, L_{\uparrow,\downarrow}) = (1, n_y, n_{\uparrow,\downarrow})$, & $(n_x, 1, n_{\uparrow,\downarrow})$		$H_1 = \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=3}\left[\sum_{\sigma\in\{\uparrow,\downarrow\}} -\left(tc^\dagger_{i,j,\sigma}c_{i,j+1,\sigma} + tc^\dagger_{i,j+1,\sigma}c_{i,j,\sigma}\right) - \frac{\mu}{2}n_{i,j,\sigma}\right] + \frac{U}{2}n_{i,j,\uparrow}n_{i,j,\downarrow}$ $H_2 = \sum_{i=1}^{n_x=4}\sum_{j=1}^{n_y=3}\left[\sum_{\sigma\in\{\uparrow,\downarrow\}} -\left(tc^\dagger_{i,j,\sigma}c_{i+1,j,\sigma} + tc^\dagger_{i+1,j,\sigma}c_{i,j,\sigma}\right) - \frac{\mu}{2}n_{i,j,\sigma}\right] + \frac{U}{2}n_{i,j,\uparrow}n_{i,j,\downarrow}$ (D121)

[1] J. Haah, A. W. Harrow, Z. Ji, X. Wu, and N. Yu, Sample-optimal tomography of quantum states, IEEE Trans. Inf. Theory **63**, 5628 (2017).

[2] S. Arunachalam and R. De Wolf, Optimal quantum sample complexity of learning algorithms, J. Mach. Learn. Res. **19**, 1 (2018).

[3] R. Takagi, H. Tajima, and M. Gu, Universal sampling lower bounds for quantum error mitigation, Phys. Rev. Lett. **131**, 210602 (2023).

[4] P. Gokhale, O. Angiuli, Y. Ding, K. Gui, T. Tomesh, M. Suchara, M. Martonosi, and F. T. Chong, Minimizing state preparations in variational quantum eigensolver by partitioning into commuting families, arXiv:1907.13623.

[5] A. Jena, S. Genin, and M. Mosca, Pauli partitioning with respect to gate sets, arXiv:1907.07859.

[6] V. Verteletskyi, T.-C. Yen, and A. F. Izmaylov, Measurement optimization in the variational quantum eigensolver using a minimum clique cover, J. Chem. Phys. **152**, 124114 (2020).

[7] W. J. Huggins, J. R. McClean, N. C. Rubin, Z. Jiang, N. Wiebe, K. B. Whaley, and R. Babbush, Efficient and noise resilient measurements for quantum chemistry on near-term quantum computers, npj Quantum Inf. **7**, 1 (2021).

[8] A. Gresch and M. Kliesch, Guaranteed efficient energy estimation of quantum many-body Hamiltonians using shadow grouping, Nat. Commun. **16**, 689 (2025).

[9] L. Nützel, A. Gresch, L. Hehn, L. Marti, R. Freund, A. Steiner, C. D. Marciniak, T. Eckstein, N. Stockinger, S. Wolf *et al.*, Solving an industrially relevant quantum chemistry problem on quantum hardware, Quantum Sci. Technol. **10**, 015066 (2025).

[10] H.-Y. Huang, R. Kueng, and J. Preskill, Predicting many properties of a quantum system from very few measurements, Nat. Phys. **16**, 1050 (2020).

[11] C. Hadfield, S. Bravyi, R. Raymond, and A. Mezzacapo, Measurements of quantum Hamiltonians with locally-biased classical shadows, Commun. Math. Phys. **391**, 951 (2022).

[12] C. Hadfield, Adaptive Pauli shadows for energy estimation, arXiv:2105.12207.

[13] A. Elben, S. T. Flammia, H.-Y. Huang, R. Kueng, J. Preskill, B. Vermersch, and P. Zoller, The randomized measurement toolbox, Nat. Rev. Phys. **5**, 9 (2022).

[14] R. Brieger, M. Heinrich, I. Roth, and M. Kliesch, Stability of classical shadows under gate-dependent noise, Phys. Rev. Lett. **134**, 090801 (2025).

[15] O. Crawford, B. V. Straaten, D. Wang, T. Parks, E. Campbell, and S. Brierley, Efficient quantum measurement of Pauli operators in the presence of finite sampling error, Quantum **5**, 385 (2021).

[16] A. Chatterjee, P. Stevenson, S. De Franceschi, A. Morello, N. P. de Leon, and F. Kuemmeth, Semiconductor qubits in practice, Nat. Rev. Phys. **3**, 157 (2021).

[17] G. Burkard, T. D. Ladd, A. Pan, J. M. Nichol, and J. R. Petta, Semiconductor spin qubits, Rev. Mod. Phys. **95**, 025003 (2023).

[18] F. Arute *et al.*, Quantum supremacy using a programmable superconducting processor, Nature **574**, 505 (2019).

[19] A. Blais, A. L. Grimsmo, S. M. Girvin, and A. Wallraff, Circuit quantum electrodynamics, Rev. Mod. Phys. **93**, 025005 (2021).
[20] F. Verstraete, J. I. Cirac, and J. I. Latorre, Quantum circuits for strongly correlated quantum systems, Phys. Rev. A **79**, 032316 (2009).
[21] A. Y. Kitaev, Quantum computations: Algorithms and error correction, Russ. Math. Surv. **52**, 1191 (1997).
[22] C. M. Dawson and M. A. Nielsen, The Solovay-Kitaev algorithm, Quantum Info. Comput. **6**, 81 (2006).
[23] G. Kuperberg, Breaking the cubic barrier in the Solovay-Kitaev algorithm, arXiv:2306.13158.
[24] H. P. Robertson, The uncertainty principle, Phys. Rev. **34**, 163 (1929).
[25] M. B. Hastings and T. Koma, Spectral gap and exponential decay of correlations, Commun. Math. Phys. **265**, 781 (2006).
[26] M. Henkel, Statistical mechanics of the 2D quantum XY model in a transverse field, J. Phys. A: Math. Gen. **17**, L795 (1984).
[27] Y. Nishiyama, Multicritical behavior of the fidelity susceptibility for the 2D quantum transverse-field XY model, Eur. Phys. J. B **92**, 1 (2019).
[28] E. Ising, PhD thesis, *Beitrag zur Theorie des Ferro-Und Paramagnestismus*, University of Hamburg (Hamburgische Universität, Hamburg, 1924).
[29] L. Onsager, Crystal statistics. I. A two-dimensional model with an order-disorder transition, Phys. Rev. **65**, 117 (1944).
[30] T. D. Schultz, D. C. Mattis, and E. H. Lieb, Two-dimensional Ising model as a soluble problem of many fermions, Rev. Mod. Phys. **36**, 856 (1964).
[31] Henk W. J Blöte and Y. Deng, Cluster Monte Carlo simulation of the transverse Ising model, Phys. Rev. E **66**, 066110 (2002).
[32] V. V. Shende, I. L. Markov, and S. S. Bullock, Minimal universal two-qubit controlled-not-based circuits, Phys. Rev. A **69**, 062321 (2004).
[33] V. V. Shende, S. S. Bullock, and I. L. Markov, in *Proceedings of the 2005 Asia and South Pacific Design Automation Conference* (Shanghai China, 2005), p. 272.
[34] S. Khatri, R. LaRose, A. Poremba, L. Cincio, A. T. Sornborger, and P. J. Coles, Quantum-assisted quantum compiling, Quantum **3**, 140 (2019).
[35] L. Madden and A. Simonetto, Best approximate quantum compiling problems, ACM Trans. Quantum Comput. **3**, 1 (2022).
[36] R. Hornreich, R. Liebmann, H. G. Schuster, and W. Selke, Lifshitz points in Ising systems, Z. Phys. B: Condens. Matter **35**, 91 (1979).
[37] T. Eckstein, R. Mansuroglu, P. Czarnik, J.-X. Zhu, M. J. Hartmann, L. Cincio, A. T. Sornborger, and Z. Holmes, Large-scale simulations of Floquet physics on near-term quantum computers, npj Quantum Inf. **10**, 84 (2024).
[38] V. Murg, F. Verstraete, and J. I. Cirac, Variational study of hard-core bosons in a two-dimensional optical lattice using projected entangled pair states, Phys. Rev. A—At., Mol. Opt. Phys. **75**, 033605 (2007).
[39] J. Jordan, R. Orús, and G. Vidal, Numerical study of the hard-core Bose-Hubbard model on an infinite square lattice, Phys. Rev. B—Condens. Matter Mater. Phys. **79**, 174515 (2009).
[40] H. A. Gersch and G. C. Knollman, Quantum cell model for bosons, Phys. Rev. **129**, 959 (1963).
[41] Y. Yanay, J. Braumüller, S. Gustavsson, W. D. Oliver, and C. Tahan, Two-dimensional hard-core Bose–Hubbard model with superconducting qubits, npj Quantum Inf. **6**, 58 (2020).
[42] N. G. Zhang and C. L. Henley, Stripes and holes in a two-dimensional model of spinless fermions or hardcore bosons, Phys. Rev. B **68**, 014506 (2003).
[43] J. Hubbard, Electron correlations in narrow energy bands, Proc. R. Soc. Lond. A. Math. Phys. Sci. **276**, 238 (1963).
[44] E. H. Lieb, Two theorems on the Hubbard model, Phys. Rev. Lett. **62**, 1201 (1989).
[45] W. Dür, M. Hein, J. I. Cirac, and H.-J. Briegel, Standard forms of noisy quantum operations via depolarization, Phys. Rev. A—At., Mol. Opt. Phys. **72**, 052326 (2005).
[46] G. González-García, R. Trivedi, and J. I. Cirac, Error propagation in NISQ devices for solving classical optimization problems, PRX Quantum **3**, 040326 (2022).
[47] Quantinuum Ltd., Quantinuum Hardware Specifications, https://www.github.com/CQCL/quantinuum-hardware-specifications
[48] A. Anshu, Concentration bounds for quantum states with finite correlation length on quantum spin lattice systems, New J. Phys. **18**, 083011 (2016).
[49] J. L. Romero, G. Björk, A. B. Klimov, and L. L. Sánchez-Soto, Structure of the sets of mutually unbiased bases for n qubits, Phys. Rev. A—At., Mol. Opt. Phys. **72**, 062310 (2005).
[50] M. Kliesch and I. Roth, Theory of quantum system certification, PRX Quantum **2**, 010201 (2021).