Efficient Classical Simulation and Benchmarking of Quantum Processes in the Weyl Basis

França, Daniel Stilck; Strelchuk, Sergii; Studziski, Micha

Published in: Physical Review Letters

DOI: 10.1103/PhysRevLett.126.210502

Publication date: 2021

Document version: Publisher's PDF, also known as Version of record

Document license: Unspecified

Efficient Classical Simulation and Benchmarking of Quantum Processes in the Weyl Basis

Daniel Stilck França,1 Sergii Strelchuk,2 and Michał Studziński3

1QMATH, Department of Mathematical Sciences, University of Copenhagen, Universitetsparken 5, 2100 Copenhagen, Denmark
2DAMTP, Centre for Mathematical Sciences, University of Cambridge, Cambridge CB30WA, United Kingdom
3Institute of Theoretical Physics and Astrophysics, National Quantum Information Centre, Faculty of Mathematics, Physics and Informatics, University of Gdańsk, Wita Stwosza 63, 80-308 Gdańsk, Poland

(Received 20 October 2020; accepted 15 April 2021; published 24 May 2021)

One of the crucial steps in building a scalable quantum computer is to identify the noise sources which lead to errors in the process of quantum evolution. Different implementations come with multiple hardware-dependent sources of noise and decoherence making the problem of their detection manyfoldly more complex. We develop a randomized benchmarking algorithm which uses Weyl unitaries to efficiently identify and learn a mixture of error models which occur during the computation. We provide an efficiently computable estimate of the overhead required to compute expectation values on outputs of the noisy circuit relying only on the locality of the interactions and no further assumptions on the circuit structure. The overhead decreases with the noise rate and this enables us to compute analytic noise bounds that imply efficient classical simulability. We apply our methods to ansatz circuits that appear in the variational quantum eigensolver and establish an upper bound on classical simulation complexity as a function of noise, identifying regimes when they become classically efficiently simulatable.

Any device designed to take advantage of quantum-mechanical features is susceptible to noise which accompanies the underlying physical realization. Reliable error correction is one of the major challenges which prevents us from building scalable hardware. The resource overhead to implement even the simplest error-correcting schemes that underpin fault-tolerant computation are currently prohibitively costly. This motivated a flurry of research into quantum algorithms [1] that work on quantum computers with small, but non-negligible, error rates and take advantage of quantum information processing protocols before the era of universal, error-corrected quantum computers. One of the key challenges is to precisely understand and characterize the noise and decoherence effects affecting these devices and to investigate how the noise affects the complexity of their classical simulation.

Successful error mitigation relies on correctly identifying the parameters of the underlying error models. The latter are constructed by employing gate-dependent benchmarking suites [2–4] which aim to characterize the singular sources of noise. One way to use the acquired knowledge about the noise in quantum computing scenarios is to introduce a quantitative measure such as quantum volume [5,6]. The latter requires us to compute the largest achievable depth of a model (random) quantum circuit that can be executed on quantum hardware by estimating the number of “heavy” output strings it generates. While this may give some insight into the reliability of the quantum computer, it has several apparent limitations.

First, it utilizes Haar-random circuits and thus does not provide the means to understand hardware performance when implementing a given quantum circuit.

Second, estimating quantum volume has an unfavorable scaling with a system size because the underlying heavy output generation problem scales exponentially with a number of qubits [7].

In our work, we introduce an approach to randomized benchmarking and classical simulation of quantum circuits that relies on Weyl unitaries. It enables us to identify a number of error models and demonstrates favorable scaling with the system size which works both for qubit and higher dimensional systems. In particular, we can identify and detect mixtures of channels such as depolarizing and dephasing channels affecting the implementation of a given gate. Having access to noiseless Clifford gates, we can also identify the parameters of other noise models including over-rotations. Moreover, our protocol is robust to the so-called state preparation and measurement errors and scalable under a natural assumption that the noise is local.

Second, we find a surprising connection between benchmarking protocols in the Weyl basis and the ability to simulate outputs of quantum circuits on a classical computer. For a given quantum circuit with the established noise profile, we provide an analytic bound on the sufficient number of samples required to classically estimate the circuit output up to a given precision using a Feynman pathlike algorithm. Thus, our methods can be used to give an upper bound on the computational power of the noisy quantum device with a clear operational
Algorithm 1. Circuit sampling.

Input.—Noisy quantum circuit specified by quantum channels \( N^{(1)}, \ldots, N^{(N)} \), initial quantum state \( \rho \) and observable \( E \).

Output.—Complex number \( x \) such that \( E(x) = \text{tr}(E\rho) \).

1. Sample \((a_0, b_0)\) from the distribution \( p_0 \).
2. For \( k = 1, \ldots, n \) sample \((a_k, b_k)\) from \( p_k(a_{k+1}, b_{k+1}|a_k, b_k) \).
3. Output \( x \) given by

\[
x = \text{sign}[(\rho(a_0, b_0))][|\rho||E(a_n, b_n)\prod_{k=1}^{N} \|N^{(k)}(a_k, b_k)\|_{\ell_1}] \text{sign}(\langle N^{(k)}(a_{k+1}, b_{k+1})\rangle).
\]

interpretation: we can establish a nontrivial computable bound on the gate noise that needs to be added to each gate in the circuit in order to render it classically efficiently simulatable. All preexisting methods of efficient classical simulation must necessarily assume a particular structure of gate set, whereas our simulation algorithm does not rely on these assumptions. Moreover, our tools do not depend on the geometry of the circuit, which provides an advantage over the state of the art tensor network methods which classically simulate quantum circuits by contracting a tensor network with cost exponential in the tree width of the graph induced by the circuit [8].

The algorithm scales particularly well for estimating the expectation value of Pauli observables on the output of local, low-depth circuits. As this is the main subroutine in quantum algorithms for near-term devices [9–11], our tools can be readily used to bound the classical simulation complexity of a wide range of quantum devices used for example in the variational quantum eigensolver (VQE) regime.

Weyl unitaries.—Our protocol makes use of Weyl-Heisenberg unitaries \( W_{(a,b)} \), which present the generalization of the Pauli matrices to higher dimensions. They are defined as \( W_{(a,b)} = Z^a X^b \), where \( X \) is the shift unitary \( X \in U(d) \), which acts on the computational basis mapping \( |j\rangle \mapsto |j+1\text{mod}d\rangle \), and \( Z \in U(d) \) is the phase unitary mapping \( |j\rangle \mapsto e^{i(2\pi j/d)}|j\rangle \), \( j = 0, \ldots, d-1 \). These unitaries have a number of useful properties: they are a (projective) representation of Weyl unitaries describing a system consisting of \( d \) qudits, which present the generalization of the Pauli matrices to higher dimensions.

Working with Weyl unitaries brings forth the importance of using compact yet rich representation space for studying quantum processes. In this setting we go beyond standard benchmarking and make use of the information about noise in the circuit to bound its classical simulation complexity. Every state \( \rho \) can be represented as a vector with respect to this basis by setting \( \rho_{(a, b)} = d^{-n/2}\text{tr}(W_{(a, b)}^\dagger \rho W_{(a, b)}) \). The same holds for observables, and quantum channels. Consider a (noisy) circuit \( C_B = N^{(N)} \circ \cdots \circ N^{(1)} \), where the \( N^{(i)}: M_{ip} \rightarrow M_{ip} \) are quantum channels, acting on a product input \( \rho = \rho_1 \otimes \cdots \otimes \rho_n \). For a given observable \( E \), we can classically simulate \( C_B \) if we can estimate \( \text{tr}(\sigma E) \) classically up to an additive error \( \epsilon > 0 \), where \( \sigma = C_B(\rho) \) (this is also known as the \textit{weak} simulation).

We will make use of the \( \ell_1 \) norm of matrices with respect to this basis, denoted by \( \| \cdot \|_{\ell_1} \) (which is just the \( \ell_1 \) norm of \( \rho \) when regarded as a vector). This is not to be confused with the usual Schatten norms.

Let \( N(a, b) \) be the \((a, b)\)th column of the channel \( N \). Then \( \|N(a, b)\|_{\ell_1} = \sum_{(a', b') \in \mathbb{Z}^2_d} |\langle N(a, b) \rangle_{(a', b')}| \), and \( \|N\|_{\ell_1} = \max_{(a, b)} \|N(a, b)\|_{\ell_1} \) is the maximum over all \((a, b)\). Our algorithm is based on \( \ell_1 \) sampling of matrices and vectors. We will assume that given quantum channels \( N^{(k)} \) in our (noisy) circuit, one can get the sample corresponding to its action on an input basis element \((a, b)\). That is, we assume we can sample from \( p_k(a, b | a', b') = |\langle N^{(k)}(a, b) \rangle_{(a', b')}|/\|N^{(k)}(a, b)\|_{\ell_1} \) and from \( p_0(a, b) = |\langle \rho(a, b) \rangle|/\|\rho\|_{\ell_1} \). As highlighted in [12], it may be convenient to work in the Heisenberg picture, which requires replacing \( \rho \) with \( E \) in the equation above. For simplicity, we present the algorithm in the Schrödinger picture.

Here \( \text{sign}(\cdot) \) function denotes the phase prefactor. Variations of this algorithm have recently and independently been discussed in other contexts [12,13]. The following theorem proves the correctness of the algorithm by showing that it samples from the true distribution:

\textbf{Theorem 1.—}The output of the circuit sampling algorithm satisfies \( E(x) = \text{tr}(\sigma E) \). Taking \( O(1/\epsilon^2 M_B \log(1/\delta)) \) many samples, where

\[
M_B = \left( \|\rho\|_{\ell_1} \|E\|_{\ell_1} \prod_{k=1}^N \|N^{(k)}\|_{\ell_1} \right)^2
\]

(1)

suffices to guarantee that with probability at least \( 1 - \delta \) an empirical average \( \tilde{x} \) of the samples satisfies \( |\tilde{x} - \text{tr}(\sigma E)| \leq \epsilon \).
1. Draw a random \((a_0, b_0) \in (\mathbb{Z}_d)^{2n}\) corresponding to the diagonal element we wish to learn and a sequence length \(m\). Initial state \(\rho\) and POVM element \(E\) on \(n\) qudits.

**Output.**—Complex number \(y\).
1. Draw a random \((a_0, b_0) \in (\mathbb{Z}_d)^{2n}\), apply \(W(a_0, b_0)\) followed by a sequence \(\hat{W} = (W(a_1, b_1), \ldots, W(a_m, b_m))\) of uniformly random local Weyl unitaries on the \(n\) qudits interspersed with the (noisy) unitary \(U\).
2. Apply \(\hat{W}^\dagger\).
3. Measure the state with a POVM \(\{E, 1 - E\}\).
4. When \(E\) is measured, output \(y = \chi_{(a, b)}(a_0, b_0) = \exp[i(2\pi/d)\langle (b - a), (a_0, b_0) \rangle]\). Else, output \(y = 0\).

**Proof.**—The probability of a fixed sequence \(((a_0, b_0), \ldots, (a_N, b_N))\) is given by
\[
\frac{|\hat{\rho}(a_0, b_0)|}{\|\hat{\rho}\|_{\ell_1}} \sum_{(a_1, b_1), \ldots, (a_N, b_N)} |\langle N^{(k)} \rangle_{(a_0, b_0)}| \prod_{k=1}^{N} |\langle N^{(k)} \rangle_{(a_k, b_k)}|^{\|E\|_{\ell_1}}.
\]
Forming a product of the probability for a given sequence in (2) and the corresponding output of the algorithm we get its expectation value by summing over all possible outcome sequences:
\[
\text{tr}(\sigma E) = \sum_{(a_1, b_1), \ldots, (a_N, b_N)} E(a_N, b_N) Z_N \cdots Z_1 \rho(a_0, b_0),
\]
where \(Z_i = \langle N^{(i)} \rangle_{(a_i, b_i)}\). The bound on the necessary number of samples follows from Hoeffding’s inequality [14] after observing that the absolute of the output in the algorithm is at most \(\|\hat{\rho}\|_{\ell_1} \sum_{k=1}^{N} |\langle N^{(k)} \rangle_{(a_k, b_k)}| \|E\|_{\ell_1} \rightarrow \chi_{(a, b)}\).

The proof only relies on the linearity of the evolution and not any property of the basis or underlying maps and vectors. Therefore, it can be easily reexpressed in the Heisenberg picture, i.e., by replacing sampling from \(\rho\) with sampling from \(E\) and \(\langle N^{(k)} \rangle^*\) in reverse order. This is useful when \(\|E\|_{\ell_1}\) is smaller than \(\|\rho\|_{\ell_1}\) or when we estimate averages of strings of Pauli operators. This sampling routine is remarkably versatile: we extend our sampling algorithm to the case of quantum circuits that made up of quantum channels of the form \(\mathcal{E}\), where \(\mathcal{L}\) is a Lindbladian. In addition, it also applies to unitary evolutions defined by Hamiltonian dynamics.

Our circuit sampling algorithm extends that of [15] in several ways. First, we show that one can use the results of the randomized benchmarking experiments to bound the complexity of a given noisy device in the Weyl basis. Second, it works for noisy quantum circuits in continuous and discrete time and evolutions in the Heisenberg picture. We present a range of bases for these as well as the Lindbladian case in Secs. VI and VII of the Supplemental Material (SM) [16], respectively.

**Weyl randomized benchmarking (WRB).**—An important feature of the Weyl operators is that many practically relevant noise models, such as (local) dephasing or (local) depolarizing channels, are diagonal in the Weyl operator basis. We will refer to such channels as Weyl diagonal channels and denote them as \(T\). It turns out that for \(d = 2\) they coincide with mixed Pauli channels and for \(d > 2\) correspond to convex combinations of conjugations with the Weyl operators; [17], Chap. 4.

Thus, given their ubiquity and the fact that randomized compiling protocols can even bring arbitrary noise to this form [18], the goal of our protocol will be to learn the parameters of a Weyl-diagonal channel that models the noise affecting a universal through randomized benchmarking protocols [19–21].

When implementing a known unitary \(U\) acting on \(n\) qudits, the resulting transformation, due to noise effects, is described by the quantum channel \(\mathcal{T} \circ \mathcal{U}\), where \(\mathcal{U}\) is the channel which corresponds to conjugation with \(U\) followed by \(T\). Our goal is to learn the parameters of a Weyl diagonal channel \(T\), i.e., its diagonal elements in the Weyl basis.

To implement the protocol we make the following assumptions about the noise: (a) one can implement Weyl unitaries with negligible error, and (b) successive implementations of \(\mathcal{U}\) are followed by the same error channel \(T\). We discuss how to relax the first assumption in Sec. II of the SM [16]. The protocol consists of the following steps:

1. The function \(\chi_{(a, b)}\) is the character of a representation of the group \((\mathbb{Z}_d)^{2n}\) and it ensures that we project the initial state to \(W_{(a, b)}\). More precisely, it follows from standard representation theory that for any operator \(Y \in \mathcal{M}_{d^2}\):
\[
(1/d^n) \sum_{(a_0, b_0) \in (\mathbb{Z}_d)^n} \chi_{(a, b)}(a_0, b_0) W_{(a_0, b_0)} Y W_{(a_0, b_0)}^\dagger = (1/d^n) \text{tr}(W_{(a, b)}^\dagger Y W_{(a, b)}).
\]

This forces the expectation value for a fixed sequence length \(m\) to be given by
\[
\frac{1}{d^n} \text{tr}[W_{(a, b)}^\dagger T \circ \mathcal{U}(W_{(a, b)})]^m S.
\]

where \(S = \text{tr}(W_{(a, b)}^\dagger \rho) \text{tr}(E W_{(a, b)})\). (See Sec. I of SM [16] for a gentle introduction of character randomized benchmarking and Sec. II for the justification of the above.) Our protocol is related to character randomized benchmarking of [3], with the distinction that we wish to determine the noise affecting a specific unitary assuming that the noise affecting Weyl operators is negligible. The protocol does not significantly depend on the particular choice of \(E\) and \(\rho\), as long as \(\text{tr}(W_{(a, b)}^\dagger \rho) \text{tr}(E W_{(a, b)}) \sim d^n\). This is because we
will later perform an exponential fit of Eq. (5) to a curve of the form $a \times e^{\mu x}$, and the magnitude of $a$ is determined by $E$ and $\rho$. Canonical choices to achieve this scaling would be to pick $E$ as the projector onto the $+1$ eigenspace of $W(\mathbf{a}, \mathbf{b})$ and $\rho$ as one of its eigenvalues because for this choice we have $\text{tr}(\hat{E}W(\mathbf{a}, \mathbf{b})) \geq d^{p-1}$ and measuring these POVMs only requires product measurements. We refer to Sec. II D of the SM [16] for more details.

By selecting different sequence lengths and performing an exponential fitting one gets an estimate of the diagonal in the Weyl basis of $\mathcal{T} \oplus \mathcal{U}$. The maximal sequence length is determined by the spectral gap $\lambda$ of the quantum channel $\mathcal{T}$. For symmetric (i.e., $\mathcal{T} = \mathcal{T}^*$) Weyl-diagonal channels, this reduces to $1 - \lambda_2$, where $\lambda_2$ is second largest eigenvalue, and is a natural measure of the noisiness of the channel. The parameter $\lambda^{-1}$ should be thought of as the depth at which the noise clearly manifests itself, as $\lambda$ can be seen as a generalized error probability of the channel. The parameter $\lambda$ can be determined using the initial knowledge for the range for noise parameters reported by the manufacturer by combining this information with a standard technique called hyperparameter grid search. For instance, for a depolarizing channel with depolarizing probability $p$, $\lambda = p$, and we expect to see errors at depth $p^{-1}$. We then have:

Theorem 2.—Let $\mathcal{T}: \mathcal{M}_{d_\mathcal{T}} \rightarrow \mathcal{M}_{d_\mathcal{T}}$ be a symmetric (i.e., $\mathcal{T} = \mathcal{T}^*$) Weyl-diagonal channel, $\mathcal{U}: \mathcal{M}_{d_\mathcal{U}} \rightarrow \mathcal{M}_{d_\mathcal{U}}$ a unitary channel, $\epsilon, \delta > 0$ be given error parameters, and let $\lambda$ be the spectral gap of $\mathcal{T} \oplus \mathcal{U}$. Then we can find an estimate $\hat{\mu}(\mathbf{a}, \mathbf{b})$ of $\mu(\mathbf{a}, \mathbf{b}) = \langle \mathcal{T} \oplus \mathcal{U} | (\mathbf{a}, \mathbf{b}) \rangle$ satisfying $|\mu(\mathbf{a}, \mathbf{b}) - \hat{\mu}(\mathbf{a}, \mathbf{b})| \leq O(\epsilon)1 - |\mu(\mathbf{a}, \mathbf{b})|^2$ with probability at least $\delta$ by performing $M = O(\epsilon^{-2}\log(\delta^{-1}\log(1 - \lambda^{-1}))$ Weyl randomized benchmarking experiments each containing at most $M_{\max} = O(\lambda^{-1})$ gates in the sequence.

Proof of Theorem 2 is located in Sec. IV of the SM [16] and is based on extending the results of [22] to our setting. We note that in the case of qubits the underlying channels are always symmetric and many other relevant examples such as dephasing or depolarizing channel belong to this class. The knowledge about $\mu(\mathbf{a}, \mathbf{b})$ and the fact that $\mathcal{T}$ is Weyl diagonal makes it sufficient to estimate $\mu(\mathbf{a}, \mathbf{b})$ to estimate the noise parameters because in this case: $\mu(\mathbf{a}, \mathbf{b}) = \langle \mathcal{U} | \langle \mathbf{a}, \mathbf{b} \rangle^{\mathcal{T}} | \mathbf{a}, \mathbf{b} \rangle$. We thus completely characterize $\mathcal{T}$ as long as the diagonal of the unitary is nonzero. It turns out that a simple enhancement of the above protocol with a suitable noiseless Clifford gate allows one to analyze noise models with any off-diagonal contributions. Consider an element $(\mathcal{T} | \langle \mathbf{a}, \mathbf{b} \rangle)$ of $\mathcal{T}$ in the Weyl basis which we want to estimate and assume we can implement a noiseless Clifford that acts as $C W(\mathbf{a}, \mathbf{b}) C^\dagger = e^{i\phi}W(\mathbf{a}, \mathbf{b})$. Such unitary always exists and can be easily identified (as long as none of the Weyl operators involved is the identity). Applying $C$ after the target unitary gate in the randomized benchmarking experiment gives access to the desired off-diagonal entry:

$$\mu(\mathbf{a}_1, \mathbf{b}_1) = \langle \mathcal{C} \mathcal{T} \mathcal{C}^\dagger | \langle \mathbf{a}_1, \mathbf{b}_1 \rangle \rangle = e^{-i\phi} \langle \mathcal{T} \mathcal{C}^\dagger | \langle \mathbf{a}_1, \mathbf{b}_1 \rangle \rangle. \quad (6)$$

The output of the algorithm is an estimate of $\langle \mathbf{a}, \mathbf{b} \rangle$ as per Theorem 2. This method can also be used to learn any number of matrix elements of $\mathcal{T} \oplus \mathcal{U}$ by conjugating it with Pauli matrices, interpolating between a constant number of learnable noise parameters analyzed in [23] and full process tomography [24,25].

To estimate Weyl-diagonal channels one requires $O(d^{2n})$ parameters, which remains practical only for small systems. However, assuming locality of the noise it is possible to learn it efficiently. For example, suppose that the unitary $\mathcal{U}$ is the product of $2^d$ gates followed by Weyl-diagonal noise acting on the same qudits. Let the noise on the first qudit be completely characterized by the diagonals with respect to Weyl operators with $(\mathbf{a}, \mathbf{b}) = (a_1, a_2, 0, \ldots, 0, b_1, b_2, 0, \ldots, 0)$. This gives a total of $O(nd^8)$ parameters to learn, rendering the protocol efficient. We discuss the “Clifford trick” of Eq. (6) as well as the extension to more complex local noise models in Secs. II C and III of the SM [16], respectively.

Theorem 2 extends the results of [26,27] in two distinct ways. First, our techniques are not qubit specific and work for systems of arbitrary dimension. Secondly, we are able to naturally incorporate gate-dependent noise (as long as we make assumptions about how it affects the Weyl operators). Thus, we relax the assumption whereby the Weyl operators are all being affected by the same, known Weyl diagonal (noise) channel.

Applications.—Local quantum circuits: When the quantum noise channels in the circuit are local and the initial state and observable are product, then the complexity of our sampling algorithm scales polynomially. To achieve this one requires efficient estimation of the transition probabilities and sampling and/or access to entries of either the state or observable in the basis. We restrict our discussion to the Weyl basis, but the argument works for any product basis.

Suppose that each quantum channel acts on at most $m = O(1)$ qudits. If we have a product basis and $N^{(k)}$ is local, then $\langle N^{(k)} | (\mathbf{a}, \mathbf{b}) \rangle = 0$ if the strings differ outside of the support of $N^{(k)}$. This is because the action of $N^{(k)}$ does not change that element of the string and, thus, the output of $(\mathbf{a}, \mathbf{b})$ remains orthogonal to the other string. Thus, given some $(\mathbf{a}_1, \mathbf{b}_1)$ as input, it suffices to only compute $\langle N^{(k)} | (\mathbf{a}_2, \mathbf{b}_2) \rangle$ that coincides with the input on the support of $N^{(k)}$ to get the elements with nonzero probability under $p(Y | (\mathbf{a}_1, \mathbf{b}_1))$. As there are only $(d^m = O(1))$ many of these, computing the associated quantities such as the normalization and signs can be done
in polynomial time, resulting in the efficient routine to which produces samples. It remains to estimate $M_B$ to determine a bound on the required number of samples. Note that

$$\|\Lambda^{(1)} \otimes \Lambda^{(2)}\|_{\ell_1-\ell_1} = \|\Lambda^{(1)}\|_{\ell_1-\ell_1} \|\Lambda^{(2)}\|_{\ell_1-\ell_1}$$ (7)

in the case of a product basis. Moreover, the $\ell_1 \to \ell_1$ norm is submultiplicative as a matrix norm induced by a vector norm, i.e., $\|\Lambda^{(1)} \otimes \Lambda^{(2)}\|_{\ell_1-\ell_1} \leq \|\Lambda^{(1)}\|_{\ell_1-\ell_1} \|\Lambda^{(2)}\|_{\ell_1-\ell_1}$.

The above properties are used for getting an estimate of $M_B$. We use the multiplicativity of the $\ell_1 \to \ell_1$ norm given by Eq. (7) for subsequences of the circuit consisting of quantum channels that do not overlap. Each individual $\ell_1 \to \ell_1$ norm can be computed efficiently and the multiplicativity implies that the overall $\ell_1 \to \ell_1$ norm of this sequence of operations is just the product of each one. Then, whenever two quantum channels have a nontrivial overlap, we may use the submultiplicativity of the norms and compute the norm for subsequences consisting of nonoverlapping quantum channels. In short, we see that if $\Lambda^{(N)} \otimes \Lambda^{(N-1)} \cdots \otimes \Lambda^{(1)}$ is a sequence of local, noisy gates that describe the circuit, then the number of samples is at most as in (1) in Theorem 1. Thus we can sample efficiently in the Weyl basis from circuits consisting only of local quantum channels.

The Weyl basis has many advantages over the phase space basis when simulating algorithms on near-term quantum hardware and studying the effects of noise. First, Clifford gates represented in this basis do not increase the sample complexity of the algorithm: they act as signed permutations in the Weyl basis and, thus, $\|\mathcal{C}\|_{\ell_1-\ell_1} = 1$ for any Clifford gate $\mathcal{C}$ and we can simulate Clifford circuits efficiently even in the presence of noise. Second, if the initial state is product and the target observable is local or is a Pauli string, then we can also achieve that $\|E[I]\|_{\ell_1} = O(1)$ by simulating the evolution in the Heisenberg picture (see Sec. V of the SM [16]).

Simulating VQE ansätze.—The simplicity of representation of Clifford gates as well as Pauli observables makes this method suitable for classically simulating quantum circuits that appear in the VQE algorithm. We apply out tools to the problem of solving MAXCUT on a graph with $n$ vertices using the VQE algorithm [28]. The problem is encoded in the ground state of the Hamiltonian $H = \sum_{i<j}^n w_{ij} \sigma^z_i \otimes \sigma^z_j$, where $w_{ij} \in \mathbb{R}$. The ansatz circuit used in this case for the state preparation has the form $|\psi(\theta)\rangle = |U(\theta)U_{\text{est}}^\dagger\rangle |\psi(\theta)\rangle$, $\theta = \{\theta_{i,k}\}_{1 \leq i \leq n, 1 \leq k \leq D}$ where the $k$th application of parametrized unitary is given by $U(\theta) = \otimes_{i=1}^n Y(\theta_{i,k})$, $Y(\theta_{i,k}) = \exp[-i(\theta_{i,k}/2)\sigma^z_i]$, and $U_{\text{est}} = \otimes_{i=1}^{n/2-1} \text{CNOT}_{2i,2i+1}$. The VQE algorithm works by iteratively preparing states $|\psi(\theta)\rangle$ which are the approximations of the ground state of $H$, where $\theta$ in each iteration are determined by a suitable classical optimization algorithm.

Now assume we performed the WRB protocol and estimated that each CNOT gate in the ansatz experiences a two-local depolarizing noise $p_C$ and the single-qubit rotations suffer from one local depolarizing noise with rate $p_Y$. We assume this rate to be independent of $\theta$ for simplicity. CNOT gate is a Clifford gate, but the $Y(\theta)$ are in general non-Clifford gates. Taking into account the noise, the process of state evolution in the Weyl basis can be represented as $\rho(\theta)_W = \Lambda^{(D)} \otimes \Lambda^{(N)} \otimes \cdots \otimes \Lambda^{(1)} |\psi(\theta)_W\rangle\langle \psi(\theta)|$. Note that both the noise and the gates are unital. To sample from the circuit we turn to Theorem 1. Clifford unitaries act as signed permutations, making $\ell_1 \to \ell_1$ norm equal to 1. Bounds are improved if we also incorporate the noise affecting the Clifford gates in the rotations: we assume that the CNOTs are noiseless, while each pair of rotations is preceded by two-local depolarizing noise $T_{p_C}$ and proceeded by one-local depolarizing noise $S_{p_Y}$. Viewing this evolution in the Pauli basis, a simple computation shows that for $T_{p_C} \otimes Y(\theta) \otimes (S_{p_Y} \otimes S_{p_Y})$ we have

$$\|T_{p_C} \otimes Y(\theta) \otimes (S_{p_Y} \otimes S_{p_Y})\|_{\ell_1-\ell_1} = \max\{1, p_C^2 p_Y \phi(\theta), p_Y p_C \phi(\theta), p_Y p_C \phi(\theta), p_Y p_{C} \phi(\theta)\},$$ (8)

where $\phi(\theta) = |\cos(\theta)| + |\sin(\theta)| \leq \sqrt{2}$. To sample the energy values with respect to $H$ which contains at most $n^2$ two-body Pauli observables with accuracy $\epsilon$ we require at most $M = O(e^{-2n^2(p_C^2 p_Y^2)^{2nD}})$ samples for $p_Y \geq 2^{-1/2}$. In particular, the algorithm is efficient whenever $p_C^2 p_Y^2 < 1/2$. Thus, as long as $2p_C^2 p_Y^2 \leq 1 + \log(nD)/nD$ we have that the number of samples scales as $M = O(n^4 D^2 e^{-2})$, making the algorithm efficient because $(p_C^2 p_Y^2)^{2nD} \leq n^2 D^2$.

D. S. F. was supported by VILLUM FONDEN via the QMATH Centre of Excellence under Grant No. 10059 and the European Research Council (Grant Agreement No. 818761). S. S. acknowledges support from the European Research Council (Grant Agreement No. 126, 210502 (2014)). The Quantum Matter and Engineering Centre is supported by the Danish National Research Foundation through the Centre for Quantum Materials and Technologies (grant No. 1271/MOB/IV/"ść