Variational Adiabatic Gauge Transformation on Real Quantum Hardware for Effective Low-Energy Hamiltonians and Accurate Diagonalization

Laura Gentini[1,2], Alessandro Cuccoli[1,2] and Leonardo Banchi[1,2,*]

[1]*Dipartimento di Fisica e Astronomia, Università di Firenze, Sesto Fiorentino (FI) I-50019, Italy*
[2]*INFN, Sezione di Firenze, Sesto Fiorentino (FI) I-50019, Italy*

Effective low-energy theories represent powerful theoretical tools to reduce the complexity in modeling interacting quantum many-particle systems. However, common theoretical methods rely on perturbation theory, which limits their applicability to weak interactions. Here we introduce the Variational Adiabatic Gauge Transformation (VAGT), a nonperturbative hybrid quantum algorithm that can use nowadays quantum computers to learn the variational parameters of the unitary circuit that brings the Hamiltonian to either its block-diagonal or full-diagonal form. If a Hamiltonian can be diagonalized via a shallow quantum circuit, then VAGT can learn the optimal parameters using a polynomial number of runs. The accuracy of VAGT is tested through numerical simulations, as well as simulations on Rigetti and IonQ quantum computers.

I. INTRODUCTION

Low-energy approximations permeate many-body physics. Systems as diverse as cold atoms in optical lattices [1], solid-state spin systems [2], and even current superconducting quantum computers [3] are accurately described by low-energy theories. Powerful theoretical methods have been developed to obtain low-energy Hamiltonians from perturbative expansions, such as the Schrieffer-Wolff transformation [4], or nonperturbative methods [5]. However, due to the exponentially large Hilbert space, numerical calculations rapidly become unfeasible when the dimensionality of the Hilbert space increases, while analytical results are limited to toy models.

Quantum computers and simulators [6,7] are starting to become experimentally available, even in the cloud [8]. It has been shown that quantum computers can accurately approximate the ground state of many-particle systems [9–11], estimate molecular energies [12], molecular docking configurations [13], and even some excited states [14]. One of the main challenges in quantum simulation is computing the dynamics of quantum many-particle systems without having to resort to exact diagonalization or conventional perturbation theory. Algorithms based on the Suzuki-Trotter decomposition [7,15] have been adapted to better exploit the capabilities of current noisy hardware [16], while accurate evolutions for longer times

can be obtained using variational fast-forwarding [17] or Hamiltonian diagonalization [18,19] methods.

Here we introduce a hybrid variational quantum algorithm for either block or full diagonalization of N-qubit Hamiltonians, where complex calculations in exponentially large Hilbert spaces are performed on a quantum hardware while, when some assumptions are met, the classical part of the algorithm scales polynomially in N. Our algorithm provides an efficient way of finding a variational circuit that brings the Hamiltonian to the desired diagonal or block-diagonal form, which can be used to extract low-energy interactions or estimate quantum dynamics as in fast forwarding. Our method is based on recent advances [20–24] in the context of adiabatic gauge potentials (AGPs), which are infinitesimal generators of a unitary transformation diagonalizing a given Hamiltonian. The AGP is nonperturbative, it can recover Schrieffer-Wolff transformation in the perturbative limit, and it is tightly connected to the Wegner Hamiltonian flow [25], also called the similarity renormalization group [26]. Some works [20–24] propose a variational approach to find an approximate AGP, and then use it to find an approximation of the diagonalizing unitary. Moreover, under some assumptions, they show that the approximate AGP leads to a unitary rotation that block diagonalizes the Hamiltonian, instead of fully diagonalizing it. However, in these works the parameters of the AGP are found using classical algorithms or, for simple models, analytical arguments, leaving the problem of exponentially large Hilbert spaces open.

For these reasons, here we introduce a different formulation of the AGP theory, called Variational Adiabatic Gauge

*Corresponding author. leonardo.banchi@unifi.it

Transformation (VAGT), where we directly parametrize the diagonalizing unitary as a shallow parametric quantum circuit. Although AGP and VAGT share the same cost function, the main result of this paper is to show that VAGT optimization can be performed via a hybrid classical-quantum algorithm, where a quantum computer is used to manipulate the many-body operators and Hamiltonians, while a classical device processes the quantum measurement outcomes to update the parameters, without dealing with exponentially large Hilbert spaces. We show that VAGT yields an accurate block diagonalization with few variational parameters (low depth) when the Hamiltonian has some energy separated or symmetry separated blocks, and full diagonalization when the number of parameters increases. Finally, the feasibility of our method on current noisy intermediate-scale quantum (NISQ) devices [6] is tested with experiments on the Rigetti Aspen-9 and IonQ 11-qubit quantum processors.

Our paper is organized as follows. In Sec. II we introduce VAGT optimization and the necessary quantum circuits and classical routines to obtain a good approximation of the optimal parameters. In Sec. III we test VAGT via numerical simulations on a classical computer, by simulating the low-energy sector of different many-body spin Hamiltonians, while in Sec. IV we present our simulations on the Rigetti Aspen-9 and IonQ 11-qubit quantum processors. Conclusions are drawn in Sec. V.

II. VARIATIONAL ADIABATIC GAUGE TRANSFORMATION

We focus on the diagonalization, or block diagonalization, of a Hamiltonian H, assuming that there is another Hamiltonian H_0 whose eigenvalues and eigenstates are known, and possibly easy to prepare on a quantum device. We thus split H as

$$H \equiv H_\lambda = H_0 + \lambda V, \tag{1}$$

where $\lambda V = H - H_0$, and λ models the strength of the correction. A good approximation of the ground state of H can be prepared thanks to the adiabatic theorem [27], by starting from the ground state $|g_0\rangle$ of H_0 and then slowly increasing the interaction strength λ. For an evolution time T, the approximate ground state is obtained as $|g\rangle = \mathcal{T} \exp\left(-i \int_0^T H_{\mu(t)} dt\right) |g_0\rangle$, where $\mu(t)$ is a function, typically linear in t, satisfying $\mu(0) = 0$ and $\mu(T) = \lambda$. Notice that we choose $\hbar = 1$, and so we do throughout this work. Such adiabatic preparation of the ground state is accurate and efficient when the ground state of H_μ is nondegenerate and well separated from the excited states for all $\mu \in [0, \lambda]$.

A generalization of the adiabatic ground-state preparation is given by the adiabatic gauge potential [20,21,28], which defines the infinitesimal generators of a unitary

transformation that allows the estimation of more eigenvalues, in some cases even performing full diagonalization—see also Appendix A for more details. Consider some infinitesimal generators A_μ for $\mu \in [0, \lambda]$, set the unitary

$$U_\mu = \mathcal{T}_\nu \exp\left(-i \int_0^\mu A_\nu d\nu\right) \iff A_\mu = i(\partial_\mu U_\mu) U_\mu^\dagger, \tag{2}$$

and the rotated Hamiltonian

$$\tilde{H}_\mu := U_\mu^\dagger H_\mu U_\mu, \tag{3}$$

with $\mathcal{T}_\nu$ denoting the ordering with respect to ν. Depending on the problem, we want to reach a diagonal or block diagonal $\tilde{H}_\mu$ at the end of the evolution, when $\mu = \lambda$ [see Fig. 1(a)]. We assume for the sake of simplicity that H_0 has two blocks, which, in the easiest case, correspond to two degenerate eigenvalues $h_\mathcal{P}$ and $h_\mathcal{Q}$ of H_0, with $h_\mathcal{P} \ll h_\mathcal{Q}$, that are possibly split by the term λV, as in Fig. 1(b)—the extension to Hamiltonians with more blocks is straightforward. In such a case, we call P and Q, respectively, the projectors on the low-energy and high-energy sectors, but in general P and Q can also be projectors on symmetry sectors of the Hamiltonian H_0. Note that, following the AGP literature [29], we supposed that subspaces P and Q are well defined; that means the strength of the perturbation λ is small enough to not mix different energy sectors arising from degenerate levels of H_0 [see Fig. 1(b)]. For block diagonalization we should impose that all the elements of the off-diagonal blocks are zero, i.e., $P\tilde{H}_\mu Q = 0$. Differentiating such an equation with respect to μ, we get

$$P U_\mu^\dagger G_\mu U_\mu Q = 0, \quad G_\mu := V + i[A_\mu, H_\mu]. \tag{4}$$

The latter equation can also be written as $[U_\mu^\dagger G_\mu U_\mu, H_0] = 0$ when $H_0 = h_\mathcal{P} P + h_\mathcal{Q} Q$ has only two degenerate eigenvalues, as in Fig. 1.

The adiabatic gauge potential is a particular choice of A_μ that satisfies the operator equation

$$[G_\mu, H_\mu] = 0, \tag{5}$$

see Appendix A for more details. Such an equation is stronger than Eq. (4) and, when exactly satisfied, the resulting $\tilde{H}_\mu$ has no off-diagonal elements. Approximations of the above exact solution were proposed in Refs. [20,21,29–32], based on a variational approximation of the A_μ, with optimal parameters obtained by variationally minimizing, on a classical computer, either $\|[G_\mu, H_\mu]\|$ or $\|G_\mu\|$, where $\|\cdot\|$ is the Hilbert-Schmidt norm. With some assumptions, such a variational approximation of the AGP is efficient in suppressing matrix elements between states that belong to different energy sectors or that are well separate in
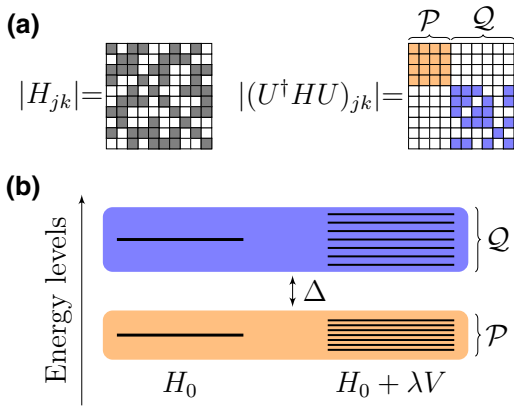
(a)

(b)

FIG. 1. Pictorial representation of the action of the adiabatic gauge transformation U, which performs a block diagonalization of $H = H_0 + \lambda V$ separating the low-energy ($\mathcal{P}$) and high-energy ($\mathcal{Q}$) Hilbert spaces. (a) Example matrix elements of H before (left) and after (right) the action of U. (b) Example structure of H_0 and V: the term V breaks the symmetries of H_0 and splits its degenerate eigenvalues, still maintaining them in two separate subspaces with suitably large energy separation Δ.

the basis defined by the symmetries of H_0, effectively resulting in a block diagonalization of the Hamiltonian. Therefore, the AGP can also be applied when low-energy and high-energy sectors are *a priori* unknown.

In this work we propose a different variational approach, whose parameters can be optimized in NISQ hardware. Before entering in the details of the quantum algorithm, we remark that the optimization procedure for AGP (see, e.g., Ref. [24]) is entirely classical. Instead, in what follows, we define a hybrid optimization procedure that finds the (block) diagonalizing unitary U_μ by taking advantage of a quantum hardware to avoid dealing with exponentially large spaces.

Taking inspiration from the success of hybrid variational quantum algorithms [33,34], we consider a variational quantum circuit ansatz for the adiabatic gauge transformation (AGT) U_μ defined in Eq. (2), namely we write

$$U_\mu(\alpha) = U_0 \prod_{\ell=1}^{L} e^{-i\alpha_\mu^\ell B^\ell}, \qquad (6)$$

where α_μ^ℓ are variational parameters, L is the number of layers in the circuit ansatz, and B^ℓ are local operators. Given the available gates in current quantum hardware we choose B^ℓ such that $e^{i\alpha B^\ell}$ is either a one- or two-qubit gate. If the Hamiltonian H_0 is diagonal in the chosen basis, then $U_0 = \mathbb{1}$, otherwise we assume that U_0 may be efficiently expressed as a known quantum circuit. Each parameter α_μ^ℓ is a continuous function of the running parameter $\mu \in [0, \lambda]$. By dividing such interval in T steps $\delta\mu$ we create a

discrete set of T values for μ:

$$\mu \in [0, \lambda] \rightarrow \{\mu_t\}_{t=1}^T, \quad \mu_t = t\delta\mu, \quad t \in \mathbb{N}. \qquad (7)$$

As a result we now have a discrete set of LT variational parameters $\{\alpha_t^\ell\}$. Within precision $\delta\mu$, defined by the choice of the number of time slice T, the potential $A_\mu = i(\partial_\mu U_\mu)U_\mu^\dagger$ at step t can be approximated via finite differences as

$$A_{\mu_t} \simeq \sum_{\ell=1}^{L} \frac{\alpha_{t+1}^\ell - \alpha_t^\ell}{\delta\mu} O_t^\ell, \qquad (8)$$

where $O_t^\ell := U_t^\ell B^\ell (U_t^\ell)^\dagger$, $U_t^\ell := U_0 \prod_{k<\ell}^{\rightarrow} e^{-i\alpha_t^k B^k}$. We set $\alpha_0^\ell = 0$ at step $t = 0$ for $1 \le \ell \le L$ and, starting from this initial configuration, we iteratively impose either Eq. (4) or (5), to get firstly α_1^ℓ and then the optimal parameters α_t^ℓ at all steps t. More precisely, as we clarify in the next sections, setting $\beta_t^\ell = (\alpha_{t+1}^\ell - \alpha_t^\ell)/\delta\mu$ those equations can be written as $X_t \cdot \beta_t = Y_t$, for some operators X_t and Y_t. Calling $\tilde{\beta}_t$ the solution of such an operator equation we get the gradientlike update rule

$$\alpha_{t+1}^\ell = \alpha_t^\ell + \tilde{\beta}_t^\ell \delta\mu, \qquad (9)$$

where all $\tilde{\beta}_t^\ell$ are obtained by classical postprocessing of quantum measurement results. We notice that, although Eq. (9) resembles a gradient-ascent update rule, it is obtained from a completely different route. Parametric quantum circuits like the one in Eq. (6) can give rise to barren plateau in the cost function landscape [35–37] when the parameters are randomly initialized or for global cost functions. However, in the VAGT algorithm the cost function is local and all the parameters are initialized to zero and then evolved to the optimal values, a strategy that has been found to address the barren plateau problem [37,38]. Some more details are provided in Appendix B. An update rule similar to Eq. (9), namely based on the solution of linear system of equations with coefficients estimated via quantum hardware, was discussed in the context of the quantum imaginary-time-evolution algorithm [33,39], but the resulting circuits are entirely different. In the following sections we study different applications that can be done efficiently on a quantum hardware.

A. Quantum circuit implementation

Since H acts on a Hilbert space whose dimension exponentially increases with the number of qubits, in general the diagonalization or block diagonalization of the Hamiltonian is exponentially hard. Here we show that, provided H can be accurately diagonalized by a shallow circuit, such diagonalizing unitary can be found in polynomial time using a hybrid quantum classical algorithm that can nowadays be run on NISQ devices. Our algorithm is based

on the minimization of the norm $\|G_\mu\|$ that, as we show in Appendix B 1, is equivalent to the solution of the linear system $\sum_\ell X_t^{l,\ell} \beta_t^\ell = b_t^l$ of L equations, from which we can update the variational parameters following Eq. (9). In order to define a quantum circuit to measure the coefficients $X_t^{l,\ell}$ and b_t^l in a quantum computer, we first expand V and H_{μ_t} in terms of Pauli operators

$$V = \sum_j v_j \sigma_j \qquad H_{\mu_t} = \sum_j h_{jt} \sigma_j, \qquad (10)$$

where $\sigma_j \equiv \sigma^{k_1^j} \otimes \cdots \otimes \sigma^{k_N^j}$ are strings of Pauli operators, $k_i^j \in \{0, x, y, z\}$ and, in principle, the sum index runs up to 4^N where N is the number of qubits. However, in most physical relevant cases the Hamiltonian contains only a limited number of terms, so most coefficients h_{jt} and v_j are null. The quantum algorithm we are about to define does not require to run a quantum circuit corresponding to such zero coefficients. In Eq. (10) we also drop the dependence on μ_t of h_j to simplify notation.

We call *computational basis* the basis in which all the Pauli operators σ_k^z are diagonal. In terms of such coefficients we find

$$b_t^l = -\sum_{j,k} v_j h_{kt} \text{Tr}(\sigma_j \, i[U_t^l B^l U_t^{l\dagger}, \sigma_k]), \qquad (11)$$

$$X_t^{l,\ell} = \sum_{j,k} h_{jt} h_{kt} \text{Tr}(i[U_t^\ell B^\ell U_t^{\ell\dagger}, \sigma_j] \, i[U_t^l B^l U_t^{l\dagger}, \sigma_k]), \quad (12)$$

where the sums are restricted to non-null values of h_{kt} and v_j. The detailed derivation of the last equations is reported in Appendix B 2, where we also show how such quantities can be estimated on a quantum computer using the circuits given in Fig. 2, where $|\phi_N\rangle$ is the maximally entangled state that can be constructed using $\mathcal{O}(N)$ operations as in Fig. 2(c). Therefore, the number of operations in each circuit is at most $\mathcal{O}(N + L)$.

The scaling efficiency of the method depends on the connectivity of the Hamiltonians H_0 and V, namely on the number of terms N_V and N_H associated to a non-null coefficient in the two quantities in Eq. (10). Suppose that $\max(N_V, N_H) = \mathcal{O}(N^\gamma)$: for instance, if the Hamiltonians contain just single-qubit terms, then $\gamma = 1$; for nearest-neighbor interactions $\gamma = 1$, too; on the other hand, $\gamma = 2$ if H or V contain all possible two-qubit interactions. For each step t, the number of circuits needed to evaluate all terms, Eqs. (11) and (12), is, respectively, $N_b = \mathcal{O}(N^{2\gamma}L)$, $N_X = \mathcal{O}(N^{2\gamma}L^2)$, where L is the number of layers in Eq. (6). Therefore, the number of measurements to be performed on the quantum device to calculate all variational parameters through Eq. (9) is

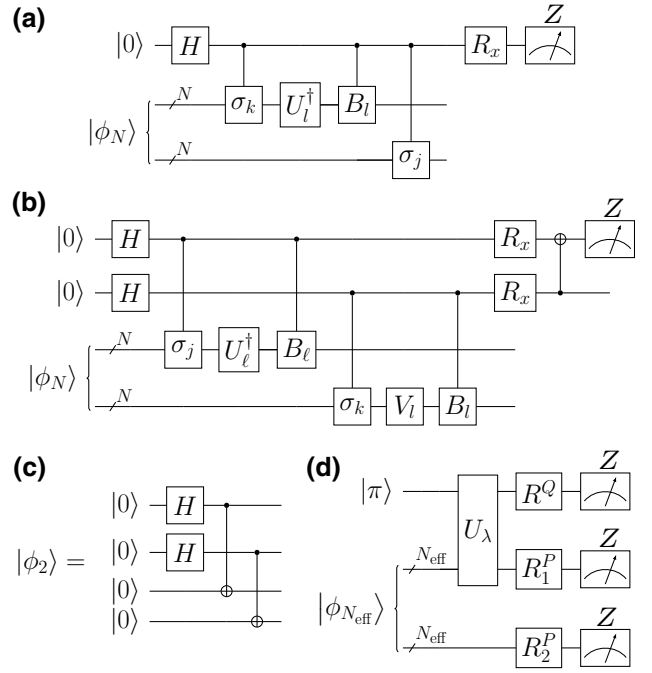$$\mathcal{O}(N^{2\gamma} L^2 T), \qquad (13)$$

FIG. 2. Suitable quantum circuits to evaluate the VAGT trace operations, appearing in Eqs. (11)–(12): (a),(b) show, respectively, the quantum circuits to measure the quantities appearing in b_t^l and $X_t^{l,\ell}$; (c) sample circuit to generate the maximally entangled input states $|\phi_N\rangle$ when $N = 2$ — see also the definition in Eq. (B23); (d) circuit to measure the expectation values in Eq. (16), where a product of single qubit rotations R_j is used to transform $\sigma_k \otimes \tilde{\sigma}_j$ into a product of Z measurements.

while the solution of all linear systems of equations is at most $\mathcal{O}(L^3)$ for each step t. Therefore, for shallow circuits with $L = \text{poly}(N)$, both the algorithmic and measurement complexities scale polynomially in the number of qubits N and linearly in the number of steps T. Finally, we remark that, when dealing with real quantum hardware, we have also to take into account the shot cost S, namely the number of times one has to repeat the same quantum measurement to evaluate an expectation value with precision that scales as $1/\sqrt{S}$.

III. NUMERICAL RESULTS

A. Low-energy approximation

When well-defined energy sectors exist for the problem at hand, as in Fig. 1, the projected block-diagonalized Hamiltonian $P\tilde{H}P$ can be interpreted as a low-energy effective Hamiltonian, that can be useful in the context of many-body physics, where the original Hamiltonian may be unmanageable for many purposes, such as calculating dynamics. Assume that P is known, and that the

low-energy block can be expanded in the Pauli basis as

$$H^{\text{eff}} = P\tilde{H}P = \sum_{\tilde{j}} h_{\tilde{j}}^{\text{eff}} \tilde{\sigma}_{\tilde{j}}, \tag{14}$$

where $\tilde{\sigma}_{\tilde{j}}$ are Pauli operators acting on N^{eff} qubits, then the expansion coefficients $h_{\tilde{j}}^{\text{eff}}$ can be obtained using a simple quantum circuit. Indeed, using the decomposition, Eq. (10), we get

$$h_{\tilde{j}}^{\text{eff}} = \sum_j \frac{h_{jT}}{2^{N^{\text{eff}}}} \text{tr}[U_\lambda^\dagger \sigma_j U_\lambda P \tilde{\sigma}_{\tilde{j}} P], \tag{15}$$

and such coefficients can be evaluated in hardware using a simple circuit like the one in Fig. 2. Indeed, suppose that $P = \mathbb{1}_{N^{\text{eff}}} \otimes |\pi\rangle_{N-N^{\text{eff}}}\langle\pi|$ and that $\tilde{\sigma}_{\tilde{j}}$ nontrivially acts only in the $\mathcal{P}$ subspace, then we may write

$$\frac{\text{tr}[U_\lambda^\dagger \sigma_j U_\lambda P \tilde{\sigma}_{\tilde{j}} P]}{2^{N_{\text{eff}}}} = \langle \pi \phi_{N_{\text{eff}}} | U_\lambda^\dagger \sigma_j U_\lambda \otimes \tilde{\sigma}_{\tilde{j}}^* | \pi \phi_{N_{\text{eff}}} \rangle, \tag{16}$$

which can be measured using the circuit shown in Fig. 2(d). In the above equations $U_\lambda \equiv U_\lambda(\alpha_{\text{opt}})$ and α_{opt} are the optimal parameters obtained at the end of the iteration, namely with $t = T$. Therefore, provided that the number N_P of non-null expansion coefficients $h_{\tilde{j}}^{\text{eff}}$ in Eq. (14) is suitably small, the effective Hamiltonian can be efficiently obtained.

We test our framework using the following three-qubit model Hamiltonian

$$H = h\sigma_3^z + \lambda\big[(\vec{\sigma}_1 \cdot \vec{\sigma}_3 + \vec{\sigma}_2 \cdot \vec{\sigma}_3) - (\sigma_1^x + \sigma_2^x)\big], \tag{17}$$

with $h = -5$. For $\lambda = 0$ the Hamiltonian is diagonal in the computational basis with only two degenerate, well-separated energy levels. As λ grows, the off-diagonal part of H is designed to completely remove the degeneracy, while keeping the levels in two separate subspaces, as showed in Fig. 3(b). In the Hamiltonian, Eq. (17), qubits 1 and 2 do not interact directly, but can effectively communicate via qubit 3. Qubit 3 is the one that define energy sectors, so we can easily identify the low-energy projector P in the computational basis as

$$P = \mathbb{1}_{12} \otimes |0\rangle_3\langle 0|, \tag{18}$$

where $|0\rangle$ is the eigenstate of σ_3^z with eigenvalue $+1$. This means that, at the end of the process, one can obtain an effective low-energy Hamiltonian that couples qubits 1 and 2, with interactions mediated by qubit 3 without having to take into account its evolution at all. Even if this is just a toy model, it is reminiscent of quantum communication schemes [40–42]: if qubit 3 is replaced by a
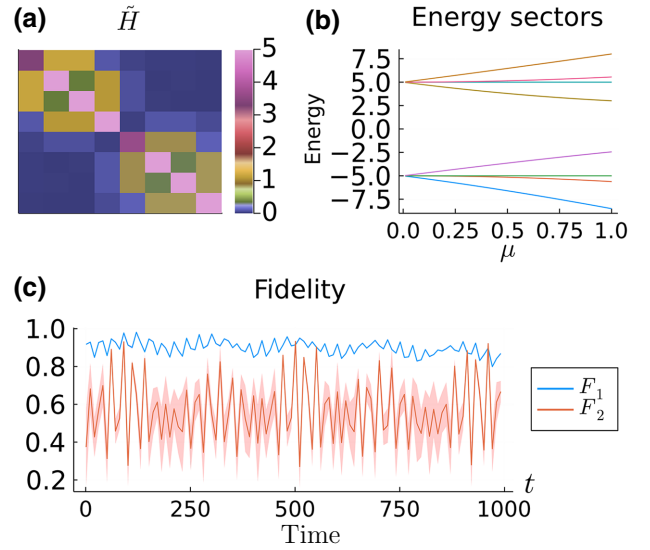

FIG. 3. Results obtained from the model Hamiltonian (17) via numerical simulation with $\lambda = 1$, $h = -5$, $T = 100$ and $L = 36$. (a) Absolute value of transformed Hamiltonian components $|\tilde{H}|_{jk}$ in the eigenbasis of H_0. (b) Energy sectors defined by the Hamiltonian (17) as μ is running from 0 to $\lambda = 1$. (c) State fidelity F_1 between the time-evolved state and the one obtained from the effective model, and fidelity F_2 between the evolved and initial states, as defined in Eq. (20), for different random initial states. Solid lines represent the mean value and colored regions the 95% confidence interval.

multiqubit communication channel, forming for example a qubit chain, then this method can be used to find an effective Hamiltonian for the sender and receiver qubits only [43].

In Fig. 3 we present the results obtained with $\lambda = 1$ and $T = 100$. We use a variational ansatz composed by two blocks of layers: the first one is made by three layers of single-qubit rotations around the x, y, and z axis, respectively, for each qubit; the second one is made of parametrized two-qubit gates, $\sigma^x \otimes \sigma^x$, $\sigma^y \otimes \sigma^y$, and $\sigma^z \otimes \sigma^z$, for different pairs of qubits. Since the Hamiltonian, Eq. (17), is symmetric with respect to the exchange of qubits 1 and 2, we employ a symmetric ansatz where each operator B^ℓ in Eq. (6) satisfies $[B^\ell, S_{12}] = 0$, being S_{12} the swap operator. Considering such symmetry and alternating and repeating each block three times, we get $L = 36$ free parameters. Figure 3(a) shows the absolute value of transformed Hamiltonian components $|\tilde{H}_{jk}|$, where the block structure due to energy sectors [Fig. 3(b)] is clearly visible. The resulting effective interaction between qubits 1 and 2 is

$$H^{\text{eff}} \simeq -1.1\,(\sigma_1^x + \sigma_2^x) + 1.0\,(\sigma_1^z + \sigma_1^z) +$$
$$- 0.2\,(\sigma_1^x\sigma_2^x + \sigma_1^y\sigma_2^y) + 0.1\,(\sigma_1^y\sigma_2^z + \sigma_1^z\sigma_2^y), \tag{19}$$

where only the terms larger than 0.1 have been shown, the full Hamiltonian can be found in Appendix C. Figure 3(c) shows the state fidelity F_1 between the time-evolved state according to the full Hamiltonian (17) and the one obtained from the effective model, together with the fidelity F_2 between the evolved and initial states

$$
\begin{aligned}
F_1(t) &= \langle \psi_{12}^{\text{eff}}(t) | \, \rho_{12}(t) \, | \psi_{12}^{\text{eff}}(t) \rangle, \\
F_2(t) &= \langle \psi_{12}^{\text{eff}}(0) | \, \rho_{12}(t) \, | \psi_{12}^{\text{eff}}(0) \rangle,
\end{aligned}
\tag{20}
$$

where $\rho_{12}(t) = \text{tr}_3 | \psi_{123}(t) \rangle \langle \psi_{123}(t) |$,

$$
| \psi_{123}(t) \rangle = e^{-itH_\lambda} | \xi_{12}, 0_3 \rangle, \quad | \psi_{12}^{\text{eff}}(t) \rangle = e^{-itH^{\text{eff}}} | \xi_{12} \rangle,
$$

$| \xi_{12} \rangle$ are randomly generated two-qubit state, and t is varied from 1 to 1000. As shown Fig 3(c), F_2 displays a nontrivial behavior, signaling a nontrivial dynamics. Since $F_1 \sim 1$ for $t \leq 1000$, such dynamics is accurately reproduced by the effective model for remarkably long times.

B. Block diagonalization

We now study the performance of the VAGT algorithm with symmetry-defined blocks, by focusing on the following spin-chain Hamiltonian with open boundary conditions

$$
H_\lambda = \sum_{i=1}^{N-1} \left(\sigma_i^x \sigma_{i+1}^x + \sigma_i^y \sigma_{i+1}^y + h \sigma_i^z \right) + \lambda \sum_{i=1}^{N} \sigma_i^x, \tag{21}
$$

where N is the number of qubits, while h and λ are, respectively, the transverse and longitudinal fields. For $\lambda = 0$ the model is exactly solvable, so eigenvalues and eigenvectors can be found in $\mathcal{O}(\text{poly}(N))$ time. For this Hamiltonian the blocks can be identified as different magnetization sectors, as H_0 commutes with $\sum_i \sigma_i^z$, without necessarily being far away in energy.

In Fig. 4 we present the results obtained with $N = 4$, $h = 4.5$, $\lambda = 1$, and $T = 100$ by numerical simulation of the quantum circuits. The VAGT in Eq. (6) is composed of two blocks of layers, the first block contains two layers of parametrized single-qubit rotation gates around the x and y axes, while the second block contains two layers of parametrized two-qubit gates. For the latter we choose only nearest-neighbor $\sigma^y \otimes \sigma^y$ and $\sigma^z \otimes \sigma^z$ interactions, and we alternate and repeat both blocks of layers 10 times, resulting in $L = 140$. Even if such an ansatz is obviously not universal for a four-qubit system, the small off-diagonal terms at the end of the optimization, as shown in the left panel of Fig. 4(a), confirm the validity of our algorithm. The solution can also be improved by using deeper circuits and finer slicing, i.e., higher T. In Fig. 4(a), we also show (right panel) the transformed Hamiltonian in the magnetization basis, where the block structure associated with the different magnetization sectors defined by the original symmetry is clearly apparent.
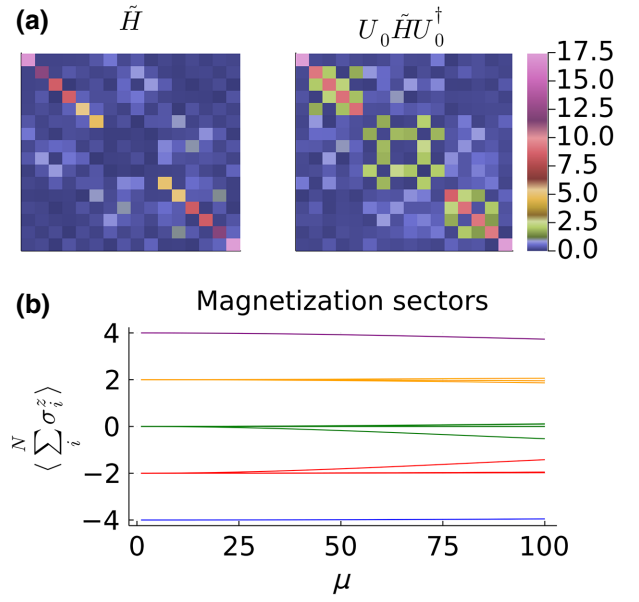

FIG. 4. (a) Results obtained from Hamiltonian, Eq. (21), for $N = 4, \lambda = 1, h = 4.5, T = 100$, and $L = 140$. The transformed Hamiltonian $\tilde{H}$ is shown both in the eigenbasis of H_0 (left) and in the magnetization eigenbasis of σ_i^z (right). (b) Magnetization sectors naturally defined by the symmetries of H_0 are split for nonzero μ. Values are obtained by analytically computing the expectation value of $\sum_i^N \sigma_i^z$ on the eigenvectors of H_μ from Eq. (21), for $\mu \in [0, \lambda]$ and $\lambda = 1$.

IV. IMPLEMENTATION ON NISQ DEVICES

We now discuss the implementation of our algorithm on real quantum hardwares, the Rigetti Aspen-9 quantum processor with 31 qubits, and IonQ quantum processor with 11 qubits, that we access through the cloud-based Amazon Braket service [8]. In order to simplify the experiment, we focus on two-qubit Hamiltonians, as in such a case, as shown in Appendix D, we can fully exploit some specific properties to minimize the number of gates employed, and accordingly the simulation cost. With this simplification, valid for $N = 2$, each circuit requires at most five qubits. In order to fully exploit Aspen-9's 31 qubits and reduce cost, we run four different experiments in parallel, still guaranteeing the presence of one or two "garbage" qubits between different experiments to reduce possible crosstalk. On the IonQ's 11-qubits hardware we run instead two experiments in parallel. In numerical simulations, the full circuits shown in Fig. 2 are implemented. Using a universal variational ansatz the circuit depth is $L = 15$, but lower depths are possible by using an ansatz suitably designed for the specific Hamiltonian problem at hand. For this purpose, we choose to test our method on quantum hardware with the highly nondiagonal Hamiltonian defined below:

$$
H_0 = \sigma_1^z + \sigma_2^z, \tag{22}
$$

$$V = v_1 \sigma_1^x + v_2 \sigma_2^x + v_3 \sigma_1^y + v_4 \sigma_2^y + v_5 \sigma_1^x \sigma_2^x +$$
$$+ v_6 \sigma_1^x \sigma_2^y + v_7 \sigma_1^y \sigma_2^x + v_8 \sigma_1^y \sigma_2^y, \qquad (23)$$

with $\lambda = 1$ and where the v_k coefficients are randomly chosen between 0 and 1. Due to the lack of well-defined sectors, whether they are defined by energy, magnetization, or other physical quantities, we do not expect block diagonalization in any basis, though thanks to the universal variational ansatz, we can expect full diagonalization in the computational basis.

In Fig. 5 we show the results of our numerical simulation and hardware experiments [44]. Figure 5(a) shows the exact energy levels for different λ and the energy levels obtained via VAGT with $T = 10$ discretization steps and $S = 100$ measurement shots. We see that, in spite of the finite discretization steps, finite measurement shots, and imperfect gate implementation, results on the IonQ hardware are very accurate, while simulations on Aspen-9 did not converge. We run different experiments on Aspen-9, always getting similar outcomes, though numerical simulations with Rigetti's decoherence and dephasing times show results comparable with IonQ. We believe that the high errors on Rigetti's hardware are possibly due to the qubit connectivity, which requires extra compilation steps in order to implement the nonlocal gates required by the VAGT circuits. On the other hand, all qubits in IonQ's hardware are fully connected, so better results are expected. Indeed, we see in Fig. 5 that the accuracy obtained with IonQ hardware is very high.

In order to study the accuracy of the computed diagonal forms, rather than focusing on operator norms or other mathematical distances, we focus on the study of physical quantities, like time-evolved correlation functions. In Fig. 6 we plot the real part of the time-evolved correlation function on the ground state:

$$C^\alpha(t) = \text{Re} \langle g_0 | U_\lambda^\dagger \sigma_1^\alpha e^{-it\tilde{H}_\lambda} \sigma_1^\alpha U_\lambda | g_0 \rangle, \qquad (24)$$

where $\alpha = x, z$, $|g_0\rangle$ is the ground state of H_0, so that $U_\lambda |g_0\rangle$ is our approximation of the ground state of the Hamiltonian; $\tilde{H}_\lambda$ is the Hamiltonian in its diagonal form, reconstructed from either experimental or simulation data. In Fig. 6 we see that all simulations provide an accurate description of the dynamics, with the exception of the simulations on the Rigetti hardware.

V. DISCUSSION AND CONCLUSION

We define the VAGT hybrid quantum algorithm for block and full diagonalization of many-body Hamiltonians. It can be used to extract low-energy effective theories in complex many-particle systems or to approximate long-time evolutions, e.g., using fast forwarding.
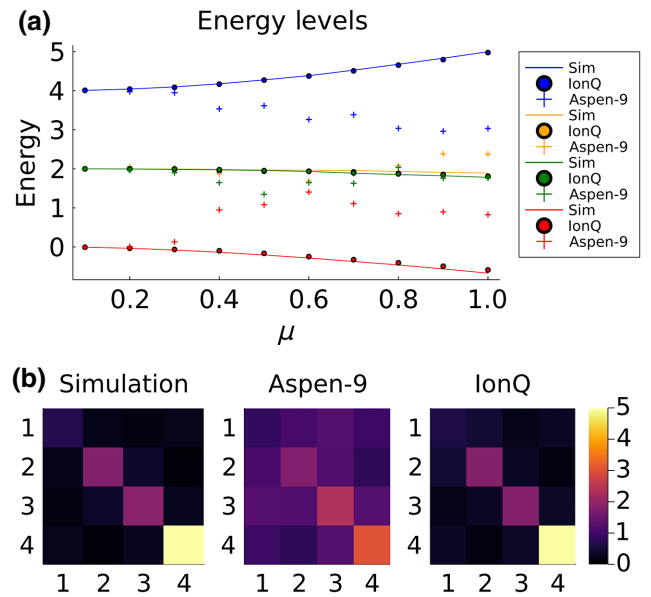

FIG. 5. (a) Energy levels for Hamiltonians, Eqs. (22) and (23), as a function of $\mu \in [0 \cdots \lambda]$ with $\lambda = 1$. (b) Absolute value of the transformed Hamiltonian components $|\tilde{H}|_{jk}$, as obtained with numerical simulations, and simulations on Rigetti's or IonQ's hardwares. All simulations are performed with $T = 10$ discretization steps and $S = 100$ shots per measurement.

The VAGT is based on the AGP, a nonperturbative method that generalizes the adiabatic theorem to multiple energy levels. The AGP has been successfully used in both analytical calculations with toy models and numerical simulations with classical computers, which are nevertheless limited to few-body operators, because of the exponentially large Hilbert space. The VAGT algorithm, on the other hand, is specifically made for hybrid quantum-classical simulations, where the complex calculations in exponentially large spaces are efficiently performed by the quantum hardware. It uses a variational quantum circuit approximation of the unitary transformation generated by the AGP, whose optimal parameters are iteratively obtained by merging outcomes from purpose-built quantum measurements with simple classical postprocessing routines. When a Hamiltonian can be transformed into a (block-) diagonal form using a shallow parametric circuit, then the VAGT algorithm can find the optimal parameters efficiently, using a number of classical and quantum operations that scale polynomially in the number of qubits.

We remind that the algorithm relays on the use of a suitable ansatz for the choice of the shallow parametric circuit: how to find such an ansatz, or even understand if it can exist for a given target Hamiltonian, are obviously rather relevant questions, which are out of the scope of the present paper.

To show the performance of the VAGT algorithm, we consider both random and physically motivated
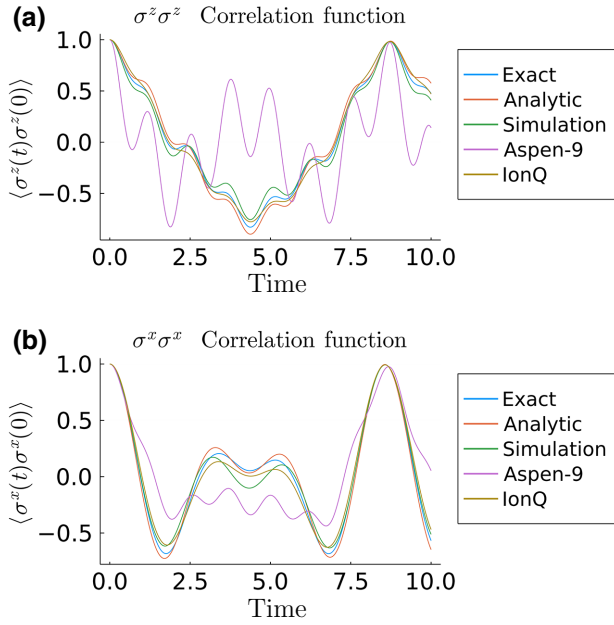
FIG. 6. Time-evolved correlation functions, as defined in Eq. (24), obtained with different methods, using the Hamiltonians, Eqs. (22)–(23). The ground states and time-evolution operators are either computed exactly (exact, analytic), through numerical simulations with shot noise (simulation), or in the quantum hardware (Aspen-9, IonQ). More precisely, in the exact line we employ analytically computed ground states of H, obtaining $|g_{\text{exact}}\rangle$ and plot $C^{\alpha}_{\text{exact}} = \text{Re} \langle g_{\text{exact}}| \sigma^{\alpha}_1 e^{-itH} \sigma^{\alpha}_1 |g_{\text{exact}}\rangle$. In all the other cases we use instead the correlation function in the form of Eq. (24), where the AGP is obtained using T steps and L layers, using the same model and parameters of Fig. 5. The analytic red line refers to a completely classical simulation of our method, where quantum measurements are replaced with analytically computed expectation values, the simulation green line refers to a classical simulation of our algorithm on an ideal quantum processor, where we implement also a quantum measurement process simulation, using $S = 100$ shots. Note that an ideal quantum processor is, by definition, unaffected by noise. The Aspen-9 and IonQ purple and brown lines show quantities computed from experimental data, obtained using real quantum hardware.

Hamiltonians, where the block structure may come from separated energy bands or may be defined by the symmetries of the problem. We perform both exact numerical simulations and simulations with realistic error sources (e.g., measurement shots), always obtaining convergence after a few iterations. Moreover, we have also run our algorithm on Rigetti and IonQ quantum computers, finding very accurate results on the latter, possibly thanks to its all-to-all qubit connectivity.

APPENDIX A: ADIABATIC GAUGE POTENTIAL

Here we briefly review the theoretical derivation of the AGP [24]. Consider a system evolving under the Hamiltonian $H_{\mu(t)}$ defined in Eq. (1), which is time dependent through the parameter $\mu(t)$ and is written as a matrix in the *computational basis* where all σ^z_j are diagonal. We define $U_{\mu(t)}$ the instantaneous unitary operator that diagonalizes the Hamiltonian at time t

$$\tilde{H}_{\mu(t)} = U^{\dagger}_{\mu(t)} H_{\mu(t)} U_{\mu(t)} \quad \forall t, \tag{A1}$$

with respect to an eigenbasis of H_0, and we use the tilde to denote operators in this reference frame. For the moving observer in the instantaneous eigenbasis of $\hat{H}_{\mu(t)}$ the effective Hamiltonian ruling the dynamics is [21]

$$\hat{H}^{\text{eff}} = \tilde{H}_{\mu(t)} - \dot{\mu} \tilde{\mathcal{A}}_{\mu(t)}, \tag{A2}$$

where $\tilde{\mathcal{A}}_{\mu(t)}$ is the AGP in the moving frame:

$$\tilde{\mathcal{A}}_{\mu(t)} = i U^{\dagger}_{\mu(t)} \partial_{\mu} U_{\mu(t)}; \tag{A3}$$

the AGP in the standard frame is

$$\mathcal{A}_{\mu(t)} = U_{\mu(t)} \tilde{\mathcal{A}}_{\mu(t)} U^{\dagger}_{\mu(t)} = i \partial_{\mu} U_{\mu} U^{\dagger}_{\mu}. \tag{A4}$$

It is possible to show [21,24] that

$$i(\partial_{\mu} H_{\mu} + F_{\text{ad}}) = [\mathcal{A}_{\mu}, H_{\mu}], \tag{A5}$$

where

$$F_{\text{ad}} = \sum_n \left(\partial_{\mu} \epsilon^n_{\mu}\right) |n_{\mu}\rangle \langle n_{\mu}| \tag{A6}$$

is the adiabatic, or generalized force, operator [21], $|n_{\mu(t)}\rangle$ being the instantaneous nth eigenstate of $H_{\mu(t)}$.

Let us now suppose we want to diagonalize the Hamiltonian $H_{\mu(t)}$ at any time t. Instead of calculating directly the unitary operator U_{μ}, we can search for its instantaneous generator, the AGP. We use a *variational* approach, in the sense that we make a hypothesis of a suitable form of $\mathcal{A}_{\mu}$, $A_{\mu}(\alpha)$, depending on some variational parameters α. Once again, here we follow the approach of Ref. [24]. Let us now

define

$$G_\mu(\alpha) = \partial_\mu H_\mu + i[A_\mu(\alpha), H_\mu]; \qquad (A7)$$

from Eq. (A5), if there is a set α^* of variational parameters such that $G_\mu = -F_{\mathrm{ad}}$ we also have

$$A_\mu(\alpha^*) = \mathcal{A}_\mu + D_\mu, \qquad (A8)$$

where D_μ is an operator that commutes with H_μ.

In Ref. [24] it is also formally demonstrated that searching for the operator A_μ, which minimizes the distance from G_μ and F_{ad}, that is searching for the best approximation of the AGP, is equivalent to finding the variational parameters that minimize $G_\mu(\alpha)$ operator norm:

$$\min_\alpha ||G_\mu(\alpha)|| = \min_\alpha ||\partial_\mu H_\mu + i[A_\mu(\alpha), H_\mu]||. \quad (A9)$$

We remind that solving this equation leads to the best approximation of the AGP, except for the diagonal part, which is undetermined by construction, as we explicitly state by Eq. (A8).

APPENDIX B: OUR VARIATIONAL ANSATZ

Taking inspiration from variational hybrid quantum classical computation [33,34], we propose to use a "quantum circuit"-type ansatz of the operator U_μ:

$$U_\mu(\alpha) = U_0 \overset{\overset{L}{\to}}{\prod_{l=1}} e^{-i\alpha_\mu^l B^l}, \qquad (B1)$$

where L is the number of layers in the circuit ansatz, the B operators are one- or two-local operators (loosely speaking, $e^{i\alpha B}$ is a one- or two-qubit gate); and the arrow over the product sign defines the order of the product itself: specifically,

$$\overset{\overset{L}{\to}}{\prod_{l=1}} \mathcal{U}_l := \mathcal{U}_1 \mathcal{U}_2 \cdots \mathcal{U}_L, \qquad (B2)$$

$$\overset{\overset{L}{\leftarrow}}{\prod_{l=1}} \mathcal{U}_l := \mathcal{U}_L \mathcal{U}_{L-1} \cdots \mathcal{U}_1. \qquad (B3)$$

Note that in this paper we suppose to know the eigenvalues and eigenstates of the H_0 Hamiltonian defined in the main text in Eq. (1), meaning we can efficiently construct the quantum circuit realizing the rotation U_0 such that

$$D_0 = U_0^\dagger H_0 U_0 \qquad (B4)$$

is diagonal, and consequently it is a constant element in the definition of our variational circuit ansatz above.

We can compute the generator of U_μ, which is our variational hypothesis for A_μ, from its definition in Eq. (A4):

$$A_\mu(\alpha) = U_0 \sum_{k=1}^{L} \left[\overset{\to}{\prod_{l<k}} e^{-i\alpha_\mu^l B^l} \frac{d\alpha_\mu^k}{d\mu} B^k \overset{\leftarrow}{\prod_{l<k}} e^{i\alpha_\mu^l B^l} \right] U_0^\dagger. \quad (B5)$$

Equations (B1) and (B5) are valid for any value of μ. If we want to (block) diagonalize the Hamiltonian in Eq. (1) we can assume that μ is a running parameter, $\mu \in 0 \cdots \lambda$, and iteratively find the generator A_μ for all points $\mu \in 0 \cdots \lambda$.

At this level, each parameter α_μ^l is a continuous function of the running parameter μ. If we now divide the interval $0 \ldots \lambda$ in T intervals $\delta\mu$, we create a discrete set of T values for μ:

$$\mu \in 0 \ldots \lambda \to \{\mu_t\}_{t=1}^{T},$$
$$\mu_t = t\delta\mu, \quad t \in \{1 \ldots T\} \in \mathbb{N}. \qquad (B6)$$

As a result, we now have a discrete set of variational parameters, $\{\alpha\}_{l,t}^{L,T}$; U_{μ_t} at the step t is expressed as a parametric evolution, depending on variational parameters α_t, and the expression of the generator at the step t is now

$$A_{\mu_t}(\alpha_t, \alpha_{t+1}) = \sum_{k=1}^{L} \frac{\alpha_{t+1}^k - \alpha_t^k}{\delta\mu} O_t^k, \qquad (B7)$$

where we use the finite-difference form for the derivative of α with respect to μ and we define

$$U_t^k := U_0 \overset{\to}{\prod_{l<k}} e^{-i\alpha_t^l B^l}, \qquad (B8)$$

$$O_t^k := U_t^k B^k (U_t^k)^\dagger. \qquad (B9)$$

Since the Hamiltonian H_0 in Eq. (1) is diagonalized by U_0, we can take advantage of the fact that $U(\mu_t = 0)|_{\{\alpha\}=0} = U_0$: this means that we know that the optimized parameters $\{\alpha\}_{l,t=0}^{L,N}$ at the step $t = 0$ are all zero.

Moreover, $A_{\mu_t}(\alpha_t, \alpha_{t+1})$ is a function of only the subset of $\{\alpha\}_{l,t}^L$ and $\{\alpha\}_{l,t+1}^L$ at the step t and $t+1$, respectively. These two facts lead to an iterative method to solve Eq. (A9):

(a) first compute Eq. (A9) at the step $t = 0$ with $\{\alpha\}_{l,t=0}^L = 0$, that we know are optimized already;

(b) this gives a function of only $\{\alpha\}_{l,t=1}^L$, the subset of parameters at $t = 1$, that can be easily optimized (see Appendix B 1 for more details);

(c) once the optimal parameters for $t = 1$ have been obtained, one has to repeat the previous two steps in order to obtain the optimal parameters for $t = 2$ and so on, until one reaches the last step $t = T$, that corresponds to $\mu = \lambda$ and the correct generator $A(\lambda)$ for the unitary operator $U(\lambda)$ is finally obtained.

1. Analytic minimization

As we said previously, at the time step t we want to optimize the cost:

$$C_{\mu_t}(\alpha_{t+1}, \alpha_t) = ||\partial_{\mu_t} H_{\mu_t} + i[A_{\mu_t}(\alpha_{t+1}, \alpha_t), H_{\mu_t}]||^2 \quad \text{(B10)}$$

with respect to α_{t+1}, since α_t's are optimized already.

Using the form, Eq. (B7), and Eq. (1) for the Hamiltonian, we find

$$
\begin{aligned}
C_{\mu_t}(\alpha_{t+1}, \alpha_t) &= \left\| V + \sum_{l=1}^{L} \frac{\alpha_{t+1}^l - \alpha_t^l}{\delta\mu} i[O_t^l, H_{\mu_t}] \right\|^2 \\
&= \text{Tr}(VV) + \frac{2}{\delta\mu} \sum_{l=1}^{L} (\alpha_{t+1}^l - \alpha_t^l)\text{Tr}(VQ_t^l) + \\
&\quad + \frac{1}{\delta\mu^2} \sum_{l,\ell}^{L} (\alpha_{t+1}^l - \alpha_t^l)(\alpha_{t+1}^\ell - \alpha_t^\ell) \\
&\quad \times \text{Tr}(Q_t^l Q_t^\ell),
\end{aligned}
\quad \text{(B11)}
$$

where we use the fact that V is Hermitian and we define

$$Q_t^l = (Q_t^l)^\dagger = i[O_t^l, H_{\mu_t}]. \quad \text{(B12)}$$

In order to minimize the cost, Eq. (B11), we can now calculate the gradient and set it to zero:

$$\frac{\partial C_{\mu_t}(\alpha_{t+1}, \alpha_t)}{\partial \alpha_{t+1}^\ell} = 0 \quad \forall \ell \in 1 \cdots L, \quad \text{(B13)}$$

$$\sum_{l=1}^{L} \frac{(\alpha_{t+1}^l - \alpha_t^l)}{\delta\mu} \text{Tr}(Q_t^\ell Q_t^l) = -\text{Tr}(VQ_l^\ell) \quad \forall \ell \in 1 \cdots L. \quad \text{(B14)}$$

Equation (B14) is a linear system that we can solve numerically: once the solution for

$$\beta_t^l \equiv \frac{(\alpha_{t+1}^l - \alpha_t^l)}{\delta\mu} \quad \text{(B15)}$$

is found, we can write the optimized parameters α_{t+1}^ℓ $\forall \ell$ as

$$\alpha_{t+1}^\ell = \alpha_t^\ell + \delta\mu\beta_t^\ell. \quad \text{(B16)}$$

2. Quantum circuits

For every step $t \in 1 \cdots T$ the method involves the solution of the linear system, Eq. (B14), so, at each step, we need to compute

$$b^l = -\text{Tr}(VQ_l), \quad \text{(B17)}$$

$$X^{\ell,l} = \text{Tr}(Q_\ell Q_l) \quad \text{(B18)}$$

$\forall$ $\ell, l, \in 1 \cdots L$, where we drop the t step index to simplify the notation.

In order to evaluate them through a quantum computer, let us suppose

$$V = \sum_j v_j \sigma_j \qquad H_\mu = \sum_j h_j \sigma_j, \quad \text{(B19)}$$

where σ_k are strings of Pauli operators $\forall$ k that forms a complete basis of SU(N) (eventually, some coefficients v_j and h_j may be zero, depending on the specific model). Our quantities become

$$b^l = -\sum_{j,k} v_j h_k \text{Tr}(\sigma_j i[U^l B^l U^{l\dagger}, \sigma_k]), \quad \text{(B20)}$$

$$X^{\ell,l} = \sum_{j,k} h_j h_k \text{Tr}(i[U^\ell B^\ell U^{\ell\dagger}, \sigma_j] i[U^l B^l U^{l\dagger}, \sigma_k]), \quad \text{(B21)}$$

where we use also Eqs. (B9) and (B12), again dropping t labels.

We are now left with the task of evaluating on a quantum computer the following two types of terms:

1. $\text{Tr}(\sigma_j i[U^l B^l U^{l\dagger}, \sigma_k])$,
2. $\text{Tr}(i[U^\ell B^\ell U^{\ell\dagger}, \sigma_j] i[U^l B^l U^{l\dagger}, \sigma_k])$.

For this aim, we consider the identity below:

$$\text{Tr}(AB) = 2^N \langle\phi| A^T \otimes B |\phi\rangle, \quad \text{(B22)}$$

where A and B are Hermitian operators acting on a 2^N-dimensional Hilbert space $\mathcal{H}$, T indicates the transpose, and $|\phi\rangle$ is the maximally entangled state defined as

$$|\phi\rangle = \frac{1}{\sqrt{2^N}} \sum_{i=0}^{2^N-1} |ii\rangle \quad \in \mathcal{H} \otimes \mathcal{H}, \quad \text{(B23)}$$

and $\{|i\rangle\}$ is a orthonormal basis for $\mathcal{H}$. Using Eq. (B22) we can now express our target terms as

1. $2^N \langle\phi| \sigma_j^T \otimes i[U^l B^l U^{l\dagger}, \sigma_k] |\phi\rangle$,
2. $2^N \langle\phi| (i[U^\ell B^\ell U^{\ell\dagger}, \sigma_j])^T \otimes i[U^l B^l U^{l\dagger}, \sigma_k] |\phi\rangle$.

The choice of employing the $2N$ maximally entangled state, Eq. (B23), in order to evaluate a trace as quantum measurement output is guided by convenience. Others

[18], use the maximally mixed state of N qubits $\rho = \mathbb{I}/2^N$: even if it may seem an easier route, thanks to the fewer qubits employed, we remark that preparing ρ usually involves preparing its purification, which happens to be the state, Eq. (B23), and tracing out half of the quantum register.

Note that σ_j are strings of Pauli operators, so $\sigma_j^T = (-1)^{n_j^y} \sigma_j$, where n_j^y is the number of σ^y in σ_j. Furthermore,

$$(i[U^\ell B^\ell U^{\ell\dagger}, \sigma_j])^T = -i[(U^{\ell\dagger})^T (B^\ell)^T (U^\ell)^T, \sigma_j^T]. \quad \text{(B24)}$$

We can choose a variational ansatz where $(B^\ell)^T \equiv B^\ell \ \forall l$ (in other words, we can use an ansatz in which σ^y appear an even number of times in each operator B^ℓ), so

$$(i[U^\ell B^\ell U^{\ell\dagger}, \sigma_j])^T = -i(-1^{n_j^y})[V^{\ell\dagger} B^\ell V^\ell, \sigma_j], \quad \text{(B25)}$$

where V is our variational ansatz U, defined in Eq. (B8), taken in the reverse order, i.e.,

$$U^k := \prod_{l<k}^{\rightarrow} e^{-i\alpha^l B^l},$$

$$V^k = (U_k)^T = \prod_{l<k}^{\leftarrow} e^{-i\alpha^l (B^l)^T} = \prod_{l<k}^{\leftarrow} e^{-i\alpha^l B^l}; \quad \text{(B26)}$$

[of course one can choose an ansatz in which B^l too is not symmetric, in this case a factor $(-1)^{n_l^y}$, where n_l^y is the number of σ^y in B_l, must be inserted in the exponential in the equation above, as well as in Eq. (B25)].

Our target terms can now be written as

1. $2^N (-1)^{n_j^y} \langle\phi| \sigma_j \otimes i[U^l B^l U^{l\dagger}, \sigma_k] |\phi\rangle$,
2. $2^N (-1)^{n_j^y+1} \langle\phi| i[V^{\ell\dagger} B^\ell V^\ell, \sigma_j] \otimes i[U^l B^l U^{l\dagger}, \sigma_k] |\phi\rangle$.

For the first type of terms, we can consider the quantum circuit of Fig. 2(a), where R_x is

$$R_x = \frac{1}{\sqrt{2}}(\mathbb{I} - i\sigma^x), \quad \text{(B27)}$$

and the construction of the state $|\phi\rangle$ from the standard initial state $|00\cdots 0\rangle$ is made by applying the Hadamard gate on the first N qubit, followed by N CNOT gates, controlled by the first N qubit with the second half of the register as a target. In particular, for $N = 2$ the quantum circuit is the one of Fig. 2(b), which constructs the state $|\phi\rangle$ and can be used as an input to the circuit above. Once the ancilla qubit is measured in the σ_z basis, the probability $p_A(0)$ of getting the outcome 0 is linearly related to the target:

$$\langle\phi| \sigma_j \otimes i[U^l B^l U^{l\dagger}, \sigma_k] |\phi\rangle = 2 - 4p_A(0). \quad \text{(B28)}$$

Similarly, the circuit of Fig. 2(c) can be employed to evaluate the second type of terms, as their values are encoded

in the probability of getting the outcome 0 from the first ancilla qubit measurement through the relationship

$$\langle\phi| i[V^{\ell\dagger} B^\ell V^\ell, \sigma_j] \otimes i[U^l B^l U^{l\dagger}, \sigma_k] |\phi\rangle = -4 + 8p_A(0). \quad \text{(B29)}$$

3. Discussion about barren plateau

The update rule, Eq. (9), resembles gradient ascent, with β playing the role of the gradient. In Ref. [35] it was shown that as the number of qubits increases it becomes more likely to start the evolution in flat areas of the parameter space where the gradient is very small, severely slowing down the optimization of the parameters. In spite of the similarities with gradient ascent, our β is an entirely different quantity, resulting from the analytic minimization of Eq. (B10). Nonetheless, it is interesting to study the strength of the β_t vector, because if β_t is very small we may end up having the same problems of gradient ascent.

To formally study this point, we derive a lower bound for the solution of the linear system of equations $X\beta = b$ (we drop the dependence on t to simplify the formulae). Let $X = U\lambda V^\dagger$ be the singular value decomposition of X, where U, V are unitary matrices and λ is diagonal with $\lambda_{ii} \equiv \lambda_i > 0$. Then, using the ℓ_2 norm we find

$$\|\beta\|^2 = \|V\lambda^{-1}U^\dagger b\|^2 = \|\lambda^{-1}U^\dagger b\|^2 =$$

$$= \sum_j \lambda_j^{-2} |(U^\dagger b)_i|^2 \geq \frac{1}{\lambda_{\max}^2} \sum_j |(U^\dagger b)_i|^2 =$$

$$= \frac{\|U^\dagger b\|^2}{\lambda_{\max}^2} = \frac{\|b\|^2}{\lambda_{\max}^2}, \quad \text{(B30)}$$

where $\lambda_{\max} = \max_j \lambda_j$. Hence $\|\beta\| \geq \|b\|/\lambda_{\max}$. In order to have an estimate of such quantities we note that $[U^\ell B^\ell U^{\ell\dagger}, \sigma_j] = \sum_k w_k^{\ell j} \sigma_k$, namely that each commutator can be expanded in the Pauli basis with σ_k different from the identity. Since $\text{tr}[\sigma_k] = 0$ the only nonzero terms in Eqs. (B20)–(B21) are those due to $\text{tr}[\sigma_k \sigma_k] = 2^N$ (see also the discussion in the above section to see how these exponential factors are treated after the measurements). Therefore, as long as $w_k^{\ell j} \neq 0$, both $\|Y\|$ and $\lambda_{\max}$ are $\mathcal{O}(2^N)$ and accordingly $\|\beta\| \gtrsim \mathcal{O}(1)$. We now study the opposite regime where $w_k^{\ell j} \simeq 0$. Indeed, barren plateau appears when the evolutions are exponentially close to their average value, namely when $U\sigma_j U^\dagger \simeq \int dU U\sigma_j U^\dagger = 0$, where dU is the Haar measure. More precisely, assuming that the U^ℓ are independent for different ℓ and using the formulae from Ref. [35] we find

$$\int dU b^\ell = 0, \quad \text{(B31)}$$

$$\int dU X^{\ell,l} = 2\delta_{l\ell} \sum_{jk} h_j h_k \text{tr}[\sigma_j \sigma_k], \quad \text{(B32)}$$

so in the worst case one might expect a small β. However, we note that the above analytic results come after several assumptions: first, it is assumed that all U^ℓ form a 2-design, and this is clearly not true since for small times the U^ℓ are close to an identity operator, since we start with $U^\ell = \mathbb{1}$; then it is assumed that the operators U^ℓ for different ℓ are independent, and also this assumption does not seem realistic. Finally, we note that the average of $\beta = X^{-1}b$ is different from $(\int dU X)^{-1} (\int dU b)$, so even when Eq. (B31) holds it is unclear if $\beta \simeq 0$. Because of all these differences, it remains an open question to formally assess whether barren plateaus may affect the performance of the VAGT algorithm. As we discuss in the main text, since our initialization is not random and the cost function is local, heuristic arguments [38] suggest that the VAGT may be free from the vanishing gradient problem.

APPENDIX C: FULL RESULTING HAMILTONIANS

The full effective Hamiltonian (19) is

$$
\begin{aligned}
H_{\text{eff}} \sim\ & -1.10\,\sigma_1^x + 0.06\,\sigma_1^y + 1.03\,\sigma_1^z \\
& - 1.10\,\sigma_2^x + 0.06\,\sigma_2^y + 1.03\,\sigma_2^z \\
& - 0.19\,\sigma_1^x \otimes \sigma_2^x - 0.04\,\sigma_1^x \otimes \sigma_2^y \\
& - 0.03\,\sigma_1^x \otimes \sigma_2^z - 0.04\,\sigma_1^y \otimes \sigma_2^x \\
& - 0.16\,\sigma_1^y \otimes \sigma_2^y + 0.10\,\sigma_1^y \otimes \sigma_2^z \\
& - 0.03\,\sigma_1^z \otimes \sigma_2^x + 0.10\,\sigma_1^z \otimes \sigma_2^y - 0.04\,\sigma_1^z \otimes \sigma_2^z,
\end{aligned}
\tag{C1}
$$

where only two decimals are significant, consistently with the choice of $T = 100$ and $\delta\mu \sim 1/T$.

APPENDIX D: SAVING MONEY: AN APPROACH FOR $N = 2$

Even if the algorithm proposed in the previous section is feasible on NISQ devices within reasonable limits, the real costs of running experiments on real devices can be high if a cost is charged for circuit reconfiguration, that is implicit at any step in our variational approach. Consequently, we develop an alternative method to reduce the number of quantum circuits employed by the algorithm itself: although this method can lead to lower costs (and we actually use it in our experiments), we remark that it is not efficient from the point of view of scalability, so that it is practically useful for very small values of N only. Dropping again the step index t, let us consider Eq. (B11):

$$
C(\alpha_{t+1}, \alpha_t) = \left\| V + \sum_{l=1}^{L} \beta_l Q_l \right\|^2,
\tag{D1}
$$

where we use definitions, Eqs. (B12) and (B15), for Q_l and β_l.

We can expand the operators V and $Q_l\ \forall l$ on the basis of Pauli strings, obtaining expressions with at most 4^N operators (this expansion is exponentially inefficient, but for $N = 2$ it leads to a 16-term expansion, that we can afford easily):

$$
V = \sum_{i=0}^{4^N} v_i \sigma_i,
\tag{D2}
$$

$$
Q_k = \sum_{j=0}^{4^N} q_{kj} \sigma_j,
\tag{D3}
$$

where σ_i is a Pauli string of two operators and, by definition

$$
y_i = \frac{1}{2^N} \mathrm{Tr}(V\sigma_i),
\tag{D4}
$$

$$
q_{kj} = \frac{1}{2^N} \mathrm{Tr}(Q_k \sigma_j).
\tag{D5}
$$

After some calculations, recalling that

$$
\mathrm{Tr}(\sigma_i \sigma_j) = 2^N \delta_{ij}
\tag{D6}
$$

we obtain

$$
C(\alpha_{t+1}, \alpha_t) = 2^N (\mathcal{V} + \mathcal{Q}\beta)^T (\mathcal{V} + \mathcal{Q}\beta),
\tag{D7}
$$

where $\mathcal{V}$ is the vector of v_i's, $\mathcal{Q}$ is the matrix composed by q_{kj} and β is the vector of β_i's. We recall that the cost function, as it is expressed in Eq. (D7) is the well-known least-squares loss function of a multiple linear regression classical problem, which can be solved efficiently.

Therefore, our goal is now to estimate efficiently the elements of $\mathcal{V}$ and $\mathcal{Q}$ defined in Eqs. (D4) and (D5), respectively.

For what concerns y_i, they are already known in our setting, since they are coefficients in Pauli decomposition of the operator V, the hard-to-diagonalize part of the Hamiltonian.

Let us focus on q_{kj}. Decomposing

$$
H_\mu = \sum_l h_l \sigma_l
\tag{D8}
$$

and taking into account Eqs. (B12) and (B22) the quantity we want to estimate is

$$
\begin{aligned}
q_{kj} &= \frac{1}{2^N} \mathrm{Tr}(Q_k \sigma_j) = \\
&= \frac{2^N}{2^N}(-1)^{n_j^y} \sum_l h_l \langle \phi | \sigma_j \otimes i[U^k B^k U^{k\dagger}, \sigma_l] | \phi \rangle,
\end{aligned}
\tag{D9}
$$

where n_j^y is the number of σ^y Pauli operators in the string σ_j. We already know that we can estimate this quantity via the first type of quantum circuit presented in Appendix B 2.
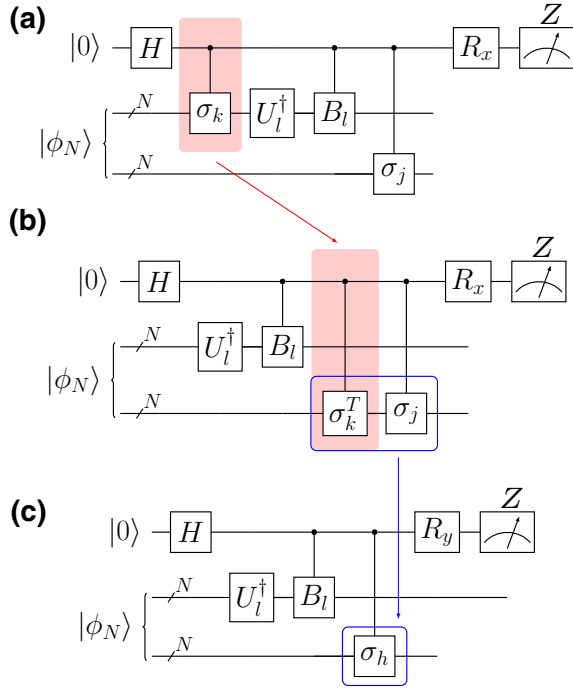
FIG. 7. Three quantum circuits carrying the same information: in (a) the original circuit employed in the general method presented in Appendix B 2; in (b) a completely equivalent circuit, with the same output; in (c) a further simplification of the circuit, where we use the information about $SU(4)$ algebra's structure coefficients.

Even for $N = 2$, the method presented above seems to perform worse than the one presented in Appendix B 2. In fact, the number of circuits we have to execute, we have

$$\mathcal{O}(N^\gamma L 4^N T), \tag{D10}$$

where N^γ is the number of terms in the decomposition, Eq. (D8), L is the length of the variational ansatz, T is the number of steps in the discretization of the parameter μ, and the highly inefficient factor 4^N comes from the decomposition of Q_k.

On the other hand, by this method we have only a quantum circuit of the type shown in Fig. 2(a), which can be reduced. In fact, consider Fig. 7: (a) shows the quantum circuit we have to execute in order to calculate q_{kj}, while (b) shows a completely equivalent quantum circuit.

The probabilities of getting the outcome 0 on the ancilla qubit are, respectively,

$$P^{ljk}_{(a)}(0) = \frac{1}{2} - \frac{1}{4} \langle\phi| \, \sigma_l \otimes i[U^k B^k U^{k\dagger}, \sigma_j] \, |\phi\rangle$$

$$P^{ljk}_{(b)}(0) = \frac{1}{2} - \frac{1}{4} \langle\phi| \, i[\sigma_l^T, \sigma_j] \otimes U^k B^k U^{k\dagger} \, |\phi\rangle \tag{D11}$$

and the equivalence between them follows from identity, Eq. (B22), and from now on we simply denote both of them with $P^{ljk}(0)$.

As we can see also from Eq. (D11) the second circuit gives a probability connected with the commutator

$$\Gamma_{lj} = [\sigma_l^T, \sigma_j] = \mathcal{F}_{ljh}\sigma_h, \tag{D12}$$

where $\mathcal{F}_{ljh}$ are the structure coefficients of the algebra. Using the last equation in Eq. (D11) we obtain

$$P^{ljk}(0) = \frac{1}{2} - \frac{1}{4}i\mathcal{F}_{ljh} \langle\phi| \, \sigma_h \otimes U^k B^k U^{k\dagger} \, |\phi\rangle . \tag{D13}$$

From the equation above it is clear that, although in principle one has to run all different $N^\gamma \cdot 4^N$ circuits in Fig. 7(b) with different Γ_{lj}, since the latter is just a commutator of two σ strings there are not so many different results for Γ_{lj}, and so there are not so many different quantum circuits one has to really run. Indeed, one can compute the structure coefficients $\mathcal{F}$ classically and run only 16 (4^N for $N = 2$) circuit of the type showed in Fig. 7(c), where σ_h is one of the 16 elements of the $SU(4)$ basis. Keeping trace of the original lj th term corresponding to a given σ_h one can recover the information about the original probability $P^{ljk}(0)$.

Note that, in the quantum circuit in Fig. 7(c) the rotation of the ancilla qubit right before the measurement process is different than that in Figs. 7(a) and 7(b):

$$R_y = \frac{1}{\sqrt{2}}\mathbb{I} - \frac{i}{\sqrt{2}}\sigma_y, \tag{D14}$$

and the output probability of getting 0 from the ancilla qubit is

$$P^{kh}_{(c)}(0) = \frac{1}{2} + \frac{1}{2} \langle\phi| \, \sigma_h \otimes U^k B^k U^{k\dagger} \, |\phi\rangle , \tag{D15}$$

and we finally recover the originally searched for probability via

$$P^{ljk}(0) = \frac{1}{2} - \frac{1}{4}i\mathcal{F}_{ljh}(2P^{kh}_{(c)} - 1). \tag{D16}$$

Finally, we remark that in the quantum circuit in Fig. 7(c) it is possible to replace the indirect measurement with a direct one: in other words, it is possible to remove the ancilla qubit, replacing the control B_k and control σ_h gates with measurements of the expectation value of $B_k \otimes \sigma_h$ on the principal register. This is possible only because both B_k and σ_h for all k and h's are Pauli strings, so they are observables.

TABLE I. Rigetti Aspen-9 calibration data.

Qubits	32
Median $T1$	33 μs
Median $T2$	16 μs
Median $1Q$ fidelity	99.39%
Median $2Q$ fidelity	94.28%

TABLE II. IonQ 11 qubit processor calibration data.

Qubits	11
Median $T1$	$> 10^7 \ \mu s$
Median $T2$	$2 \times 10^5 \ \mu s$
Median $1Q$ fidelity	99.5%
Median $2Q$ fidelity	97.5%

APPENDIX E: CALIBRATION DATA

We use real quantum processors for this work. We report in the tables calibration data for both hardwares: Rigetti Aspen-9 (Table I) and IonQ 11 qubit processor (Table II). Data are available at the Rigetti website [45] and in Ref. [46] as well as in website [47] for IonQ.

[1] L.-M. Duan, E. Demler, and M. D. Lukin, Controlling Spin Exchange Interactions of Ultracold Atoms in Optical Lattices, Phys. Rev. Lett. **91**, 090402 (2003).

[2] M. Wagner, *Unitary Transformations in Solid State Physics* (North Holland, Amsterdam and New York, 1986).

[3] P. Krantz, M. Kjaergaard, F. Yan, T. P. Orlando, S. Gustavsson, and W. D. Oliver, A quantum engineer's guide to superconducting qubits, Appl. Phys. Rev. **6**, 021318 (2019).

[4] S. Bravyi, D. P. DiVincenzo, and D. Loss, Schrieffer–Wolff transformation for quantum many-body systems, Ann. Phys. (N. Y) **326**, 2793 (2011).

[5] D. Burgarth, P. Facchi, H. Nakazato, S. Pascazio, and K. Yuasa, Eternal adiabaticity in quantum evolution, Phys. Rev. A **103**, 032214 (2021).

[6] J. Preskill, Quantum computing in the NISQ era and beyond, Quantum **2**, 79 (2018).

[7] I. M. Georgescu, S. Ashhab, and F. Nori, Quantum simulation, Rev. Mod. Phys. **86**, 153 (2014).

[8] Amazon Web Services, Amazon Braket (2021), https://aws.amazon.com/braket/, last accessed on 30-06-2021.

[9] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'brien, A variational eigenvalue solver on a photonic quantum processor, Nat. Commun. **5**, 1 (2014).

[10] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets, Nature **549**, 242 (2017).

[11] M. Motta, C. Sun, A. T. Tan, M. J. O'Rourke, E. Ye, A. J. Minnich, F. G. Brandão, and G. K.-L. Chan, Determining eigenstates and thermal states on a quantum computer using quantum imaginary time evolution, Nat. Phys. **16**, 205 (2020).

[12] P. J. O'Malley *et al.*, Scalable Quantum Simulation of Molecular Energies, Phys. Rev. X **6**, 031007 (2016).

[13] L. Banchi, M. Fingerhuth, T. Babej, C. Ing, and J. M. Arrazola, Molecular docking with Gaussian boson sampling, Sci. Adv. **6**, eaax1950 (2020).

[14] O. Higgott, D. Wang, and S. Brierley, Variational quantum computation of excited states, Quantum **3**, 156 (2019).

[15] B. Şahinoğlu and R. D. Somma, Hamiltonian simulation in the low-energy subspace, Npj Quantum Inf. **7**, 1 (2021).

[16] Y. Li and S. C. Benjamin, Efficient Variational Quantum Simulator Incorporating Active Error Minimization, Phys. Rev. X **7**, 021050 (2017).

[17] C. Cirstoiu, Z. Holmes, J. Iosue, L. Cincio, P. J. Coles, and A. Sornborger, Variational fast forwarding for quantum simulation beyond the coherence time, Npj Quantum Inf. **6**, 1 (2020).

[18] B. Commeau, M. Cerezo, Z. Holmes, L. Cincio, P. Coles, and A. Sornborger, in *APS March Meeting Abstracts*, APS Meeting Abstracts, Vol. 2021 (2021), p. X34.003.

[19] T. Jones, S. Endo, S. McArdle, X. Yuan, and S. C. Benjamin, Variational quantum algorithms for discovering Hamiltonian spectra, Phys. Rev. A **99**, 062304 (2019).

[20] J. Wurtz, P. W. Claeys, and A. Polkovnikov, Variational Schrieffer-Wolff transformations for quantum many-body dynamics, Phys. Rev. B **101**, 014302 (2020).

[21] D. Sels and A. Polkovnikov, Minimizing irreversible losses in quantum systems by local counterdiabatic driving, Proc. Natl. Acad. Sci. **114**, E3909 (2017).

[22] T. Hatomura and K. Takahashi, Controlling and exploring quantum systems by algebraic expression of adiabatic gauge potential, Phys. Rev. A **103**, 012220 (2021).

[23] S. Sugiura, P. W. Claeys, A. Dymarsky, and A. Polkovnikov, Adiabatic landscape and optimal paths in ergodic systems, Phys. Rev. Res. **3**, 013102 (2021).

[24] M. Kolodrubetz, D. Sels, P. Mehta, and A. Polkovnikov, Geometry and non-adiabatic response in quantum and classical systems, Phys. Rep. **697**, 1 (2017).

[25] F. Wegner, Flow-equations for Hamiltonians, Ann. Phys. **506**, 77 (1994).

[26] E. Anderson, S. Bogner, R. Furnstahl, E. Jurgenson, R. Perry, and A. Schwenk, Block diagonalization using similarity renormalization group flow equations, Phys. Rev., C. Nucl. Phys. **77**, 037001 (2008).

[27] T. Albash and D. A. Lidar, Adiabatic quantum computation, Rev. Mod. Phys. **90**, 015002 (2018).

[28] P. W. Claeys, M. Pandey, D. Sels, and A. Polkovnikov, Floquet-Engineering Counterdiabatic Protocols in Quantum Many-Body Systems, Phys. Rev. Lett. **123**, 090602 (2019).

[29] J. Wurtz and A. Polkovnikov, Emergent conservation laws and nonthermal states in the mixed-field Ising model, Phys. Rev. B **101**, 195138 (2020).

[30] H. Saberi, T. c. v. Opatrný, K. Mølmer, and A. del Campo, Adiabatic tracking of quantum many-body dynamics, Phys. Rev. A **90**, 060301 (2014).

[31] A. Hartmann and W. Lechner, Rapid counter-diabatic sweeps in lattice gauge adiabatic quantum computing, New J. Phys. **21**, 043025 (2019).

[32] G. Passarelli, V. Cataudella, R. Fazio, and P. Lucignano, Counterdiabatic driving in the quantum annealing of the p-spin model: A variational approach, Phys. Rev. Res. **2**, 013283 (2020).

[33] X. Yuan, S. Endo, Q. Zhao, Y. Li, and S. C. Benjamin, Theory of variational quantum simulation, Quantum **3**, 191 (2019).

[34] L. Gentini, A. Cuccoli, S. Pirandola, P. Verrucchi, and L. Banchi, Noise-resilient variational hybrid

quantum-classical optimization, Phys. Rev. A **102**, 052414 (2020).

[35] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Barren plateaus in quantum neural network training landscapes, Nat. Commun. **9**, 1 (2018).

[36] A. Arrasmith, M. Cerezo, P. Czarnik, L. Cincio, and P. J. Coles, Effect of barren plateaus on gradient-free optimization, Quantum **5**, 558 (2021).

[37] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles, Cost function dependent barren plateaus in shallow parametrized quantum circuits, Nat. Commun. **12**, 1791 (2021).

[38] E. Grant, L. Wossnig, M. Ostaszewski, and M. Benedetti, An initialization strategy for addressing barren plateaus in parametrized quantum circuits, Quantum **3**, 214 (2019).

[39] S. McArdle, T. Jones, S. Endo, Y. Li, S. C. Benjamin, and X. Yuan, Variational ansatz-based quantum simulation of imaginary time evolution, Npj Quantum Inf. **5**, 1 (2019).

[40] S. Bose, Quantum Communication Through an Unmodulated Spin Chain, Phys. Rev. Lett. **91**, 207901 (2003).

[41] L. Banchi, T. J. G. Apollaro, A. Cuccoli, R. Vaia, and P. Verrucchi, Optimal dynamics for quantum-state and entan-

glement transfer through homogeneous quantum systems, Phys. Rev. A **82**, 052321 (2010).

[42] L. Banchi, J. Fernández-Rossier, C. F. Hirjibehedin, and S. Bose, Gating Classical Information Flow via Equilibrium Quantum Phase Transitions, Phys. Rev. Lett. **118**, 147203 (2017).

[43] A. Wojcik, T. Łuczak, P. Kurzyński, A. Grudka, T. Gdala, and M. Bednarska, Multiuser quantum communication networks, Phys. Rev. A **75**, 022330 (2007).

[44] Experiments on Rigetti Aspen-9 were executed from June 15–21, 2021, while experiments on IonQ hardware were performed from August 29–September 2, 2021.

[45] Rigetti website, www.rigetti.com/what-we-build.

[46] K. Wright *et al.*, Benchmarking an 11-qubit quantum computer, Nat. Commun. **10**, 1 (2019).

[47] https://docs.microsoft.com/en-us/azure/quantum/provider-ionq, this website refers to Microsoft cloud partner for IonQ. As the IonQ team state in their own website https://ionq.com/technology, updated reliable calibration data are provided by their cloud partners like Microsoft.