Perturbative gadgets for gate-based quantum computing:
Nonrecursive constructions without subspace restrictions

Simon Cichy [1,2,*], Paul K. Faehrmann [1,*], Sumeet Khatri [1], and Jens Eisert [1,3,4]

[1]*Dahlem Center for Complex Quantum Systems, Freie Universität Berlin, 14195 Berlin, Germany*
[2]*Institute for Theoretical Physics, ETH Zürich, 8093 Zürich, Switzerland*
[3]*Helmholtz-Zentrum Berlin für Materialien und Energie, Hahn-Meitner-Platz 1, 14109 Berlin, Germany*
[4]*Fraunhofer Heinrich Hertz Institute, 10587 Berlin, Germany*

Perturbative gadgets are a tool to encode part of a Hamiltonian, usually the low-energy subspace, into a different Hamiltonian with favorable properties, for instance, reduced locality. Many constructions of perturbative gadgets have been proposed over the years. Still, all of them are restricted in some ways: Either they apply to some specific classes of Hamiltonians, they involve recursion to reduce locality, or they are limited to studying time evolution under the gadget Hamiltonian, e.g., in the context of adiabatic quantum computing, and thus involve subspace restrictions. In this work, we fill the gap by introducing a versatile universal, nonrecursive, nonadiabatic perturbative gadget construction without subspace restrictions, that encodes an arbitrary many-body Hamiltonian into the low-energy subspace of a three-body Hamiltonian and is therefore applicable to gate-based quantum computing. Our construction requires rk additional qubits for a k-body Hamiltonian comprising r terms. Besides a specific gadget construction, we also provide a recipe for constructing similar gadgets, which can be tailored to different properties, which we discuss.

I. INTRODUCTION

The study of many-body Hamiltonians is a well-researched field in condensed-matter physics and quantum information theory. While the presence of Hamiltonian terms acting on many qubits simultaneously can result in exciting phenomena, in many situations, it also carries additional difficulties. Be it due to the hardness of generating them experimentally or other limitations, local Hamiltonians are usually simpler to deal with.

Born in the context of complexity theory for proving the QMA-completeness of the local Hamiltonian problem [1,2], a problem located at the interface of the theory of quantum many-body physics and Hamiltonian complexity, so-called *perturbative gadgets* can be used to reduce the locality of a given many-body Hamiltonian. This is done by embedding it in the low-energy subspace of a tailored, local, i.e., few-body, Hamiltonian acting on a larger Hilbert space [3,4]. Among the best-known constructions is the subdivision gadget proposed by Oliveira and Terhal [4]. It employs mediator qubits to halve the support of a given Hamiltonian term and can be recursively applied to construct an at best three-body Hamiltonian mimicking the original k-body Hamiltonian. As is typical for perturbative gadgets, such a procedure, unfortunately, leads to an exponential increase in the interaction strengths required for the gadget Hamiltonian to accurately mimic the low-energy subspace of the original Hamiltonian.

A different gadget, proposed by Jordan and Farhi [5], can avoid the need for recursion and offers a direct reduction from

k-body to two-body Hamiltonians, albeit with an exponential suppression of energies. However, their direct construction is only applicable when one can restrict the evolution of chosen initial states to predefined subspaces of the Hilbert space, as can be done in adiabatic quantum computing, where commutation relations guarantee the system remains in the initialized subspace.

Since such a limitation restricts the applicability of this direct k-to-two-local gadget, the question arises whether there exists a direct, nonrecursive k-to-k'-local gadget without such restrictions and whether such a gadget could have any applications besides being a tool for proving complexity theoretic results. After all, perturbative gadgets have so far mostly found their application in proving hardness results for the general local Hamiltonian problem [3,4] or certain classes of Hamiltonians [6–10] equipped with further constraints and properties.

In this work, we address these questions by proposing a direct k-to-3-local gadget without the need for subspace restrictions, that is inspired by Ref. [5] but that avoids initialization and commutation properties from adiabatic quantum computing to achieve the same locality reduction as the repeated subdivision procedure [4], without requiring recursive applications.

As expected of gadget constructions [11], our construction is also not free from unfavorable energy scalings, but we still explore the use of such a gadget within gate-based quantum computing and, in particular, as a method for reducing the locality of the cost function in variational quantum algorithms, a setting that does not comply with the "adiabatic" restriction, or the grouping of measurement terms.

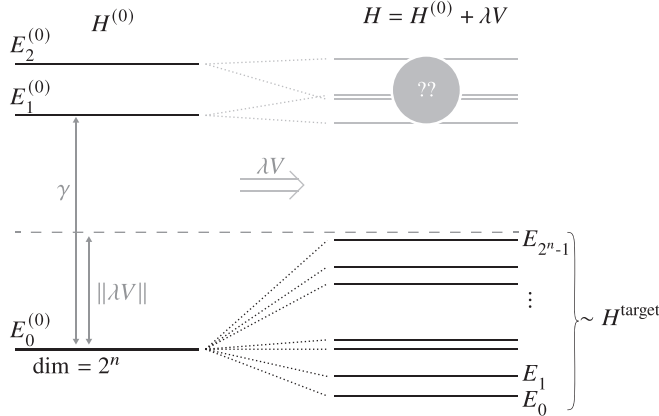*These authors contributed equally to this work.

FIG. 1. Sketch of the main idea behind perturbative gadgets. Starting from an unperturbed Hamiltonian $H^{(0)}$ with gap γ and degenerate ground space, add a perturbation V such that the lowest energy states resulting from the perturbation of the ground space mimic the target Hamiltonian H^{target}.

While most likely unfavorable for situations where the locality of a Hamiltonian scales with the system size, there are plenty of situations where the locality is constant but still high enough to cause problems. Examples include measurement noise scaling with the locality of the measurement operator influencing noise-induced barren plateaus [12] or time evolution and quantum phase estimation for restricted hardware settings. The proposed methods do not rely on subspace restrictions and can thus be implemented in gate-based quantum computing. By providing a more general recipe for constructing perturbative gadgets with similar performance guarantees and suggesting plausible heuristics, we hope to start a discussion about how the trade-offs of perturbative gadgets could still be used for practical problems.

We begin by introducing the main ideas and a historical overview of perturbative gadgets in Sec. II, before focusing on a specific k-to-3-local gadget and discussing guarantees on its performance in Sec. III. We then generalize the gadget construction to provide a recipe for creating custom perturbative gadgets in Sec. IV and discuss the nuances of applying such gadgets in gate-based quantum computing in Sec. V, before concluding with some remarks and open questions in Sec. VI.

II. PERTURBATIVE GADGETS IN A NUTSHELL

Perturbative gadgets [3–5,20] encode the low-energy subspace of a many-body Hamiltonian into the low-energy subspace of a fewer-body Hamiltonian. They make use of perturbation theory in the opposite direction than what is more common: instead of considering the effect of some perturbation λV on a well-understood Hamiltonian H according to $H' = H + \lambda V$, perturbative gadgets provide a suitable H and perturbation V such that H' is a few-body Hamiltonian that approximates the low-energy subspace of a given, target Hamiltonian H^{target}, as visualized in Fig. 1. They are often accompanied by an increase in the number of required qubits and either suppress the energy of the desired low-energy subspace or an increase in the norm of H' compared with H^{target}.

In most cases, a local and simple Hamiltonian with degenerate ground space is used as the so-called "unperturbed Hamiltonian" H. The degeneracy is chosen to correspond to the dimension of the space acted upon by the target Hamiltonian. Then, a perturbation is designed to split said degeneracy in a way that the resulting low-energy subspace emulates the target spectrum, so roughly

$$\Pi H' \Pi \sim H^{\text{target}}, \tag{1}$$

where Π denotes a projector on the low-energy subspace. This statement will later be made formal in Theorem 1.

To be explicit and clear, we use *interaction weight* to refer to the number of particles acted upon by a given operator and *interaction strength* to refer to the coefficient of that operator. When talking about qubits and expanding operators in the Pauli basis, the *weight* of a Pauli string (a tensor product of Pauli operators) is defined as the number of nontrivial Pauli operations in the string. Throughout this work, we use "locality" to refer to the maximum weight of all terms in a Hamiltonian, independently of geometry. We use the term *local* for Hamiltonians with fixed small weights, that is weights that do not scale with problem size, while *global* is reserved for Hamiltonians with all-to-all interactions.

An overview of works related to perturbative gadgets is presented in Table I. Historically, perturbative gadgets were developed to prove the QMA-hardness of the local Hamiltonian problem. The 3-local Hamiltonian problem is known to be QMA-complete [2] and Kempe, Kitaev, and Regev have shown that so is the 2-local problem by proposing the first perturbative gadget, which reduces the locality from 3 to 2 [3]. Oliviera and Terhal have added to this by showing that the 2-local Hamiltonian problem is still QMA-complete when restricting the interaction to a two-dimensional (2D) square lattice, and in doing so introduced a new perturbative gadget. This subdivision gadget reduces the locality of a k-local Hamiltonian to $(\lceil k/2 \rceil + 1)$-local interactions by introducing a single auxiliary qubit per interaction term. These kinds of gadgets have also been considered to be used outside of the world of complexity-theoretic analysis, but they are inherently not practical, as discussed further in Sec. V.

Furthermore, the recursive use of gadgets to arrive at two-local Hamiltonians can be avoided by a single gadget introduced by Jordan and Farhi in Ref. [5], which encodes k-local Hamiltonians in kth order of perturbation theory of a two-local gadget Hamiltonian. However, just like the three-to-two local gadget, their construction relies on certain properties of the evolution under a Hamiltonian where the state is restricted to a predetermined subspace, such as provided in the adiabatic setting.

In the following years, several works proposed improvements in resource requirements, convergence bounds, or coupling strengths [15–17,19], as well as introduced highly specialized gadget constructions for restricted classes of Hamiltonians, such as for topological quantum codes and for realizing certain parent Hamiltonians featuring intrinsic topological order [20,21].

There are two important things to note at this point: While all of these gadget constructions rely on perturbation theory, they use different techniques, ranging from perturbative expansion due to Bloch [22], which underlies the Jordan-Farhi

TABLE I. Overview of some relevant works related to perturbative gadgets. Not included are those that apply only to a restricted subset of Hamiltonians, such as topological Hamiltonians.

Authors	Main statement
Kempe, Kitaev, Regev (KKR) [3]	3- to 2-local gadget to prove QMA completeness of the local Hamiltonian problem. Requires restriction of the state evolution to a subspace of the Hilbert space.
Oliveira, Terhal (OT) [4]	Subdivision gadget (k- to $k/2$-local) and proof of QMA completeness of the Hamiltonian problem on a square grid. Has to be applied recursively resulting in exponential coupling strengths.
Biamonte, Love [13]	"Realizable Hamiltonians" with a focus on using gadgets for adiabatic computing and adapting to hardware restrictions, not targeting a reduction of locality.
Bravyi, DiVicenzo, Loss, Terhal [14]	Improvements on the OT gadget through extension to non-converging perturbative expansion, thus reducing the gap in coupling strengths
Jordan, Farhi [5]	Direct k- to 2-local gadget through higher-order perturbation theory. Requires restriction of the state evolution to a subspace of the Hilbert space.
Cao, Babbush, Biamonte, Kais [15]	Resource requirement improvements of some of the OT and KKR previous gadgets.
Cao [16]	Overview of existing gadgets and improvements of the OT and KKR gadgets.
Cao, Kais [17]	Improvement of the convergence bound for the JF gadget.
Subasi, Jarzynski [18]	Nonperturbative gadgets for adiabatic Hamiltonian evolution.
Bausch [19]	Augmentation of other gadget constructions and changes their energy scales by (potentially) increasing their locality by one.
Harley, Datta, Klausen, Bluhm, França, Werner, Christandl [11]	A general framework for analog quantum simulation and unavoidable unfavorable size-dependent scalings of gadget constructions.
This work	Direct k- to 3-local gadget without restriction to some subspace of the Hilbert space.

gadget, the Feynman-Dyson series employed in Ref. [6], to the Schrieffer-Wolf transformation [23], discussed in Ref. [17]. Furthermore, except for the so-called subdivision gadget due to Oliveira and Terhal [4], the other gadgets are restricted to the setting of time evolution under the Hamiltonian, which guarantees the evolution to remain in the initialized subspaces, rendering them inapplicable for gate-based quantum computing. A nonrecursive k-to-two-local gadget without these restrictions is currently unknown.

III. A DIRECT k-TO-THREE-BODY GADGET WITHOUT SUBSPACE RESTRICTION

One of the main results of this work is the proposal of a direct, i.e., nonrecursive, three-body gadget Hamiltonian H^{gad} derived from an arbitrary k-body Hamiltonian H^{target} that does not require any restrictions to a subspace of the Hilbert space, inspired by the work of Jordan and Farhi [5]. We then find that the low-energy subspace of H^{gad} mimics the low-energy subspace of H^{target} (Theorem 1), implying a guarantee on the closeness of their ground states (Corollary 1).

We start by defining the specific gadget Hamiltonian that is the focus of our study.

Definition 1 (gadget Hamiltonian). Let H^{target} be a k-body Hamiltonian acting on n qubits, given by

$$H^{\text{target}} := \sum_{s=1}^{r} c_s h_s, \quad \text{with } h_s = \sigma_{s,1} \otimes \sigma_{s,2} \otimes \cdots \otimes \sigma_{s,k}, \quad (2)$$

where each term is a tensor product of at most $k \leqslant n$ single qubit operators, with $\sigma_{s,j} = n_X X + n_Y Y + n_Z Z$ and $(n_X, n_Y, n_Z) \in \mathbb{R}^3$ a unit vector. We define the three-body gadget Hamiltonian H^{gad} corresponding to H^{target}, acting on

$n + rk$ qubits, as

$$H^{\text{gad}} := \sum_{s=1}^{r} H_s^{\text{aux}} + \lambda \sum_{s=1}^{r} V_s, \quad (3)$$

where

$$H_s^{\text{aux}} := \sum_{j=1}^{k} \frac{1}{2}\big(\mathbb{1}_{s,j} - Z_{s,j}^{\text{aux}}\big) = \sum_{j=1}^{k} |1\rangle\langle 1|_{s,j}^{\text{aux}}, \quad (4)$$

$$V_s := \sum_{j=1}^{k} \tilde{c}_{s,j} \sigma_{s,j}^{\text{target}} \otimes X_{s,j}^{\text{aux}} \otimes X_{s,(j+1) \bmod k}^{\text{aux}}, \quad (5)$$

and $\tilde{c}_{s,j} = -(-1)^k c_s$ if $j = 1$ or $\tilde{c}_{s,j} = 1$ otherwise. We refer to the n qubits on which H^{target} acts as *target qubits*, and the rk additional qubits as *auxiliary qubits*.

In our gadget Hamiltonian, each term H_s^{aux} acts only on the sth group of auxiliary qubits, while each term V_s acts on a target qubit and the sth group of auxiliary qubits; see Fig. 2 for a graphical depiction for a toy model.

The proposed gadget is heavily inspired by the gadget introduced by Jordan and Farhi in Ref. [5], in which they rely on properties of adiabatic quantum computing. Both their gadget and ours take advantage of a larger Hilbert space to encode the low-energy subspace of the target Hamiltonian, so the dimension of the ground space of the unperturbed Hamiltonian must be the same as the space on which the target Hamiltonian acts. This point is made more explicit in Appendix B. Jordan and Farhi start from a larger ground space of the unperturbed Hamiltonian and reduce the dimension by restricting the region of the Hilbert space that can be explored by initializing the system in a fixed subspace. Remaining in this chosen subspace is then ensured by using properties of adiabatic evolution, something that cannot be done straightforwardly outside of the regime of adiabatic quantum computing. This
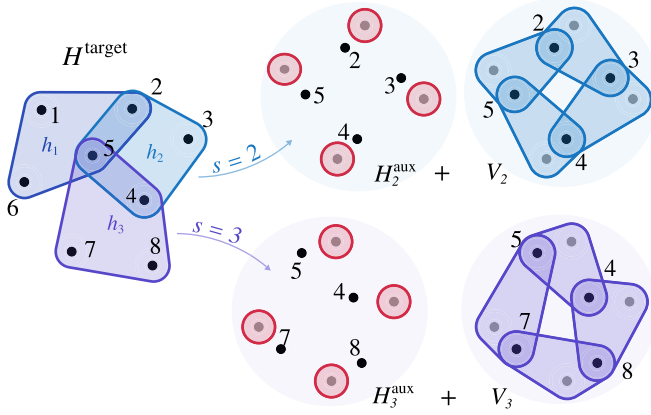
FIG. 2. Representation of a target Hamiltonian (left) and its associated gadget Hamiltonian, as defined in Eq. (3), for the case of $n = 8$, $k = 4$, and $r = 3$. Qubits that interact according to the terms of the Hamiltonian are placed in the same colored region. Shown in the central and right-hand panels are the contributions to the gadget Hamiltonian from two of the three four-body terms, with the auxiliary qubits in gray and the target qubits in black.

requirement, therefore, places a strong limitation on the applicability of their construction. On the other hand, our gadget construction does not require such a restriction to a subspace, therefore avoids this limitation, and can be applied even in the context of digital, gate-based quantum computing. We refer to Appendix F for a more detailed discussion of why a direct application of the gadget introduced by Jordan and Farhi is not possible in the realm of nonadiabatic quantum computing.

Similar to the results of Jordan and Farhi, we can show that the subspace of the 2^n lowest energy eigenstates of our three-body gadget Hamiltonian in Eq. (3) mimics that of the k-body target Hamiltonian. For the formal statement, we define the following notation: Let H be a Hamiltonian acting on n qubits, with spectral decomposition $H = \sum_{j=0}^{2^n-1} E_j |\psi_j\rangle\langle\psi_j|$ such that $E_0 \leqslant E_1 \leqslant \cdots \leqslant E_{2^n-1}$, we define $H_{\text{eff}}(H, d) = \sum_{j=0}^{d-1} E_j |\psi_j\rangle\langle\psi_j|$ to be the effective Hamiltonian corresponding to the d lowest eigenvalues of H.

Theorem 1 (main result). Let H^{target} and H^{gad} be as in Definition 1, and let $\lambda \leqslant \lambda_{\max}$, with $\lambda_{\max} = \frac{1}{4}[\sum_{s=1}^{r} |c_s| + r(k-1)]^{-1}$. Then, there exists an $f(\lambda) = O(\text{poly}\,\lambda)$ and $\Xi = O(\text{poly}\,k)$ such that

$$
\begin{aligned}
H_{\text{eff}}(H^{\text{gad}}, 2^n) &= \frac{\lambda^k}{\Xi} H^{\text{target}} \otimes (|0\rangle\langle 0|)^{\otimes rk} \\
&\quad + f(\lambda)\Pi + O(\lambda^{k+1}),
\end{aligned}
\tag{6}
$$

where Π is the projector onto the support of $H_{\text{eff}}(H^{\text{gad}}, 2^n)$.

Furthermore, we can provide a guarantee on the closeness of the ground states of the target and gadget Hamiltonians.

Corollary 1 (guarantees on the closeness of the ground states). Let H^{target} and H^{gad} be as in Theorem 1. Then, there exists a perturbation strength λ^*, with

$$
\lambda^* \leqslant \lambda_{\max} \quad \text{and} \quad \lambda^* \leqslant \frac{E_1^{\text{target}} - E_0^{\text{target}}}{\Xi \|O^{\text{err}}\|},
\tag{7}
$$

such that for all $\lambda \leqslant \lambda^*$ it holds that

$$
\|\psi_0 - \text{Tr}_{\text{aux}}[\phi_0]\|_2 \leqslant O(\lambda),
\tag{8}
$$

where $\psi_0 = |\psi_0\rangle\langle\psi_0|$ and $\phi_0 = |\phi_0\rangle\langle\phi_0|$ are in the ground spaces of H^{target} and H^{gad}, respectively.

In Fig. 3, we provide a visualization of the main steps in the proof of Theorem 1, focusing on the main ideas behind these results and referring the mathematically interested reader to the Appendixes. Specifically, after a recap in Appendix A of the perturbation theory on which our results are based, we prove Theorem 1 in Appendix B and Corollary 1 in Appendix C.

For clarity, the construction is stated for target Hamiltonians where each term has the same Pauli weight, and for the maximal reduction: to a 3-local gadget Hamiltonian. This gadget can be generalized in both of these aspects and we show how to construct a gadget for mixed Pauli weights in Appendix D 1 and also how to trade off target locality for reduced resources in Appendix D 2.

IV. A RECIPE FOR CREATING YOUR OWN PERTURBATIVE GADGET

In the previous section, we presented a specific perturbative gadget construction. However, that construction is by no means unique, because there is some freedom in designing such gadgets. For instance, the choice of Pauli operators in the unperturbed Hamiltonian acting on the auxiliary qubits or in the perturbation could have been different. More generally, the general construction of the penalization in H_s^{aux} and of the perturbation V_s leaves many knobs to tweak. Especially for practical implementations, it might be interesting to derive slightly different gadgets, which result in similar low-energy spectra but are constructed from a different set of operators, for instance, some that are easier to implement on the hardware of choice.

Here, we present a more general framework for the construction of nonrecursive, nonadiabatic perturbative gadgets based on higher-order perturbation theory that all lead to similar results as Theorem 1, with the only notable difference being the exact form of Ξ in Eq. (6).

The core idea is again to start from a well-known, unperturbed Hamiltonian H^{aux} and to perturb it with a perturbation λV, such that we recover the original, k-body target Hamiltonian at kth order in perturbation theory. Since the kth order in perturbation theory comes with k-fold applications of the perturbation V, we engineer the perturbation such that only cross terms leading to the desired result contribute. That is if the target Hamiltonian has been cut into k pieces $O^{\text{target}} = O_1 \otimes \cdots \otimes O_k$, we ensure that only the correct k-fold products of V survive.

To construct a perturbative gadget following a similar recipe as the one presented in this work, we require a *penalization Hamiltonian H* and a *perturbation operator A*, both acting on a single auxiliary register of k qubits. Although those are not the unperturbed Hamiltonian H^{aux} and the perturbation V directly, they are closely related. H^{aux} will be built using instances of H, while the perturbation operator will be a key component in the construction of the perturbation V. We further require H^{aux} to be composed of few-body terms and to
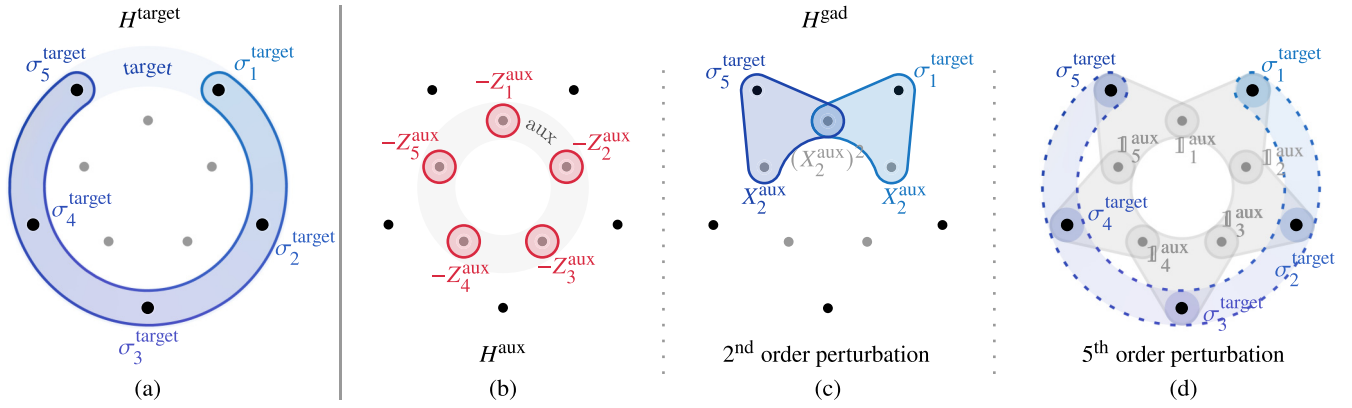
FIG. 3. Visualization of the key steps in the proof of Theorem 1. (a) For illustrative purposes, we take as the k-body target Hamiltonian $H^{\text{target}} = \bigotimes_{j=1}^{5} \sigma_j$, where $\sigma_j \in \{X, Y, Z\}$ are Pauli operators, so that $k = 5$. (b) The unperturbed auxiliary Hamiltonian, given in Eq. (4), acts only on the auxiliary qubits. (c) Acting with the perturbation V in Eq. (5), at second order, one term could result in flipping the second and last auxiliary qubits, hence ejecting the state out of the ground space and not contributing to the perturbative expansion. (d) At kth order in perturbation theory, we obtain a nontrivial contribution acting on all target qubits simultaneously, mimicking the target Hamiltonian, while acting trivially on the auxiliary qubits.

have a nondegenerate ground space to avoid the limitations of prior gadget constructions [3,5,18]. That is, we need

$$H = \sum_i h_i, \qquad (9)$$

for some finite sum of operators h_i, each of which is few-body, such that H has a unique ground-state vector, which we denote by $|GS\rangle$. We note that the restriction on few-body terms is not strictly required for the construction to yield the desired low-energy subspace, but for decreasing the locality of the Hamiltonian.

Let us now argue that the operator A should exhibit a product form of k operators $a_1, a_2, \ldots, a_k$, such that

$$A = \prod_{j=1}^{k} a_j, \qquad (10)$$

with the following properties:

$$A|GS\rangle \propto |GS\rangle, \qquad (11)$$

$$\langle GS| \prod_{\ell=1}^{m} a_{j_\ell} |GS\rangle = 0 \; \forall \, m < k \; \forall \, \{j_\ell\} \subset \{1, 2, \ldots, k\}, \qquad (12)$$

$$a_j^2 = \mathbb{1} \; \forall \, j \in \{1, 2, \ldots, k\}. \qquad (13)$$

The first condition states that the ground state of H should also be an eigenstate of A, while the second enforces that any partial application of the operators a_j results in a state orthogonal to $|GS\rangle$. These are vital to ensure that at orders lower than k in perturbation theory, the perturbation will always expel the state out of the ground space. The final property ensures that when the same perturbation is applied twice, it results in a constant energy shift.

With the above properties, we can construct H^{aux} and V as

$$H_s^{\text{aux}} = \underbrace{\mathbb{1}^{\otimes n}}_{\text{target register}} \otimes \underbrace{\mathbb{1}^{\otimes (s-1)k} \otimes H \otimes \mathbb{1}^{\otimes (r-s)k}}_{\text{auxiliary registers}}, \qquad (14)$$

$$V_s = \sum_{j=1}^{k} \tilde{c}_{s,j} \sigma_{s,j} \otimes \mathbb{1}^{\otimes (s-1)k} \otimes a_j \otimes \mathbb{1}^{\otimes (r-s)k}, \qquad (15)$$

where the definition of the coefficients $\tilde{c}_{s,j}$ provides another knob to tweak. They could be defined similarly to Definition 1 or in an arbitrary fashion, as long as

$$\prod_{j=1}^{k} \tilde{c}_{s,j} = -(-1)^k c_s \qquad (16)$$

still holds. Given these conditions, similar results as in Theorem 1 will hold due to the arguments laid out in Appendix B. Indeed, the gadget we propose is simply a special case of

$$h_i = \tfrac{1}{2}(\mathbb{1} - Z_i) = |1\rangle\langle 1|_i, \quad i \in \{1, 2, \ldots, k\}, \qquad (17)$$

$$a_j = X_j \otimes X_{(j+1) \bmod k}, \quad j \in \{1, 2, \ldots, k\}. \qquad (18)$$

Note that the Hamiltonian in Eq. (14) has a nondegenerate ground space when considering only the auxiliary registers but has a 2^n degenerate ground space when additionally considering the target qubits which it acts trivially upon.

Furthermore, we would like to stress that any gadget constructed in this fashion will be accompanied by a suppression of the target Hamiltonian in the low-energy subspace by a factor of λ^k, therefore not avoiding a blow-up in resource costs exponential in the reduction in the locality. Since the perturbative expansions only hold as long as the perturbation parameter $\lambda \leqslant \frac{1}{4}[\sum_{s=1}^{r} |c_s| + r(k-1)]$, this trade-off is ingrained into the construction and cannot be circumvented by rescaling the coefficients c_s or similar.

V. ON THE TRADE-OFFS IN USING PERTURBATIVE GADGETS FOR PRACTICAL APPLICATIONS

Perturbative gadgets have been a useful tool in complexity theory, in particular in the study of the QMA-completeness of the local Hamiltonian problem [3,4]. These successes and the locality-reducing property encourage us to look for practical applications of these kinds of gadgets outside of the world of complexity-theoretic analyses. The first candidate is adiabatic computing and analog quantum simulation [6,11,13,14]. Due to the limitations of experimental setups to implement

few-body couplings, these simulations could benefit from the locality reduction to effectively study many-body Hamiltonians while physically implementing few-body interactions. Unfortunately, the large differences in coupling strengths pose a challenge for experimental realizations [11,14,19]. Having lifted the restriction to subspaces in the gadget construction limiting previous proposals, we can now turn towards gate-based quantum computing in the search for potential applications, an area that was previously not readily accessible. In the following, we look at four specific settings that could benefit from perturbative gadgets and discuss the accompanying trade-offs, which are usually unfavorable. While we set the stage for the potential application of our gadgets to these settings, we leave the detailed study of these applications to future work.

We start with variational quantum algorithms and discuss the issue of cost-function-dependent barren plateaus and the readout problem, before noting the impact of the locality of the measurement operator on measurement noise. Afterward, we discuss possible applications in phase estimation and time evolution for restricted hardware settings.

Cost-function-dependent barren plateaus. Variational *quantum algorithms* (VQAs) [24–26] are strong in the run to find useful applications of quantum computers in the *noisy intermediate scale quantum* (NISQ) era [27] and are being intensively studied. These algorithms rely on *parametrized quantum circuits* (PQCs) to evaluate a parameter-dependent cost function related to the expectation value of a set of observables. An optimization algorithm, e.g., gradient descent, implemented on a classical device is used to optimize the parameters of the PQC to find the minimum of the cost function.

Two of the most prominent examples of VQAs might be the variational quantum eigensolver [28], aimed at finding energies of Hamiltonians, and the quantum approximate optimization algorithm [29]. One hurdle to overcome for such algorithms to become useful is the so-called barren plateau issue [12,30–33]. This refers to the phenomenon where the cost-function landscape becomes essentially flat, rendering optimization algorithms ineffective due to the lack of efficiently measurable gradients. Many strategies for mitigating barren plateaus have been proposed [32,34–43], but the observation most appealing to us is the fact that, given two cost functions with the same minimum but different localities, the one with reduced locality performs better [44–46]. This dependence of the emergence of barren plateaus on the locality of the cost function was formalized for some ansätze in Refs. [47,48].

Looking back at the results from Sec. III, our gadget enables a generic procedure for any cost function written as the expectation value of a Hamiltonian to find an equivalent 3-local cost function. This new cost function has the same minimum (Corollary 1) as the target cost function and has provably nonvanishing gradients (due to Ref. [48, Theorem 2]).

Upon first inspection, the problem of cost-function-locality-dependent barren plateaus seems to be solved, but one needs to take a step back and remember the original problem. Barren plateaus are an issue because exponentially vanishing gradients require an exponential number of measurements on a quantum device to resolve the gradients. This experimental cost means that such procedures are not scalable. Although on the surface our proposal produces gradients that can be computed efficiently, finding the minimum of the *effective* gadget Hamiltonian might become hard. Indeed, let us consider the extreme case of a global cost function, i.e., an *n*-local cost function. In this setting, from Theorem 1, the contribution of the target Hamiltonian is suppressed as λ^n. In the end, the features of interest from the target Hamiltonian might become exponentially small. Consequently, although the true minimum of the gadget cost function coincides with the minimum of the target cost function, the actual optimization may converge very close to the minimum of the gadget, but owing to the exponential suppression the energy of the target may still be very high.

Not all is lost, and we show some successful simulations in Appendix G. Our gadget-induced cost function could be used as an initialization strategy. Indeed, even if the target Hamiltonian is exponentially suppressed in the effective low-energy subspace, each of the terms of the perturbation in gadget Hamiltonian contains some part of the target Hamiltonian. It could be that using our gadget provides a useful path to reach a region of the target cost function that is outside of the regime plagued by barren plateaus. Furthermore, previous works on perturbative gadgets argue that restricting to the regime where the perturbative expansion converges is not necessarily justified [14]. In our case, that means that it could be beneficial to choose a perturbation parameter λ larger than the bound for which our analytical results hold, thus increasing the contribution of the target Hamiltonian. This idea is supported by the results of our numerical simulations. Interpreting this perturbation factor as a hyperparameter of the model, one could take inspiration from classical machine learning and implement a schedule of slowly decreasing λ, having strong contributions at the beginning and high precision at the end. These heuristics will however need to be confirmed by larger experiments than those we could perform or by further theoretical studies. An extensive description of the problem of cost-function-dependent barren plateaus and of the attempt to use perturbative gadgets, complemented with numerical simulations, can be found in Appendix E.

Reducing measurement bases for readout. Another, albeit more pessimistic example of possible trade-offs using perturbative gadgets is the construction of a gadget aimed at exploiting the commutation structure of the gadget Hamiltonian.

As for the gadget introduced in Eq. (3), by first measuring all auxiliary qubits in the Pauli-*Z* basis, we can estimate H^{aux}. Then, consecutively measuring all auxiliary qubits in the Pauli-*X* basis and simultaneously all target qubits in one of the three Pauli bases, we can estimate all of the V_s terms. Consequently, only four different measurements are required to obtain an unbiased estimator of the energy of the gadget Hamiltonian.

Going one step further, we could design a new gadget that decomposes each term of the Hamiltonian into its Pauli-*X*, Pauli-*Y*, and Pauli-*Z* parts and use each part in a single perturbation term. That would lead to a gadget that reconstructs the original term in the third order of perturbation theory and thereby reduces the complexity of the gadget.

In case the locality is of no importance, we can also reduce the number of terms in the gadget Hamiltonian and the qubit overhead. Through a different definition of the part of the perturbation in the gadget Hamiltonian that acts on the auxiliary register, we can achieve only a logarithmic qubit overhead, as shown in more detail in Appendix H.

However, we still end up facing the same challenge as before: While we can efficiently estimate the energy of the gadget Hamiltonian, that is not the quantity we are interested in. Instead, we want to measure the energy of the target Hamiltonian, which is suppressed by λ^3 and will, due to the implicit dependence of λ on the number of terms in the target Hamiltonian, lead to an unfavorable scheme compared with simply measuring each term of the Hamiltonian by itself. The only potential use case for such a gadget would then be a setting where it is excessively more expensive to change the measurement setting than to evaluate the circuit.

Measurement noise. Another setting in which the locality of certain operators can noticeably impact the performance of a gate-based quantum computer is measurement noise. As shown in the context of noise-induced barren plateaus [12], measurement noise scales directly with the locality of the measurement operator. Therefore, even when the locality of the cost function is constant in the number of qubits, even constant reductions in locality might improve the performance of the (variational) algorithm. It is important to note that this reduction will still be accompanied by energy suppression, so whether such a trade-off is favorable again highly dependent on the actual setting. However, it is an application within gate-based quantum computing, that could previously not be addressed by perturbative gadgets due to the subspace restrictions of previous constructions or their recursive nature overcomplicating implementation.

Phase estimation. The final example of a situation that could benefit from reductions in locality is that of quantum phase estimation. There, controlled time evolutions for some Hamiltonian H are combined with the quantum Fourier transform to estimate the energies of quantum states. Since approximating the time evolution, e.g., by product formulas requires the application of quantum gates whose locality scales with the locality of the Hamiltonian, the feasibility of such algorithms depends on the used hardware platform. Appropriate perturbative gadgets could then be used to circumvent limitations in hardware connectivity or reduce the number of required entangling gates. While the accompanying trade-off in the energy scale would result in improved required accuracy of its readout and consequently larger circuit depths and auxiliary qubit counts, they might be favorable in certain settings. Also, since quantum phase estimation requires the preparation of an initial state with a large overlap of the true ground state (if the goal is to estimate the ground-state energy), using perturbative gadgets does not introduce additional complications, because the previous initial state can just be extended by initializing all auxiliary qubits of the gadget in the zero state.

VI. SUMMARY AND OUTLOOK

Perturbative gadgets are a powerful tool that can be used to show equivalences between Hamiltonians of different localities. They have led to advances in quantum complexity theory, addressing problems that originate from studies of quantum many-body systems in rigorous condensed-matter theory, and have been considered tools to facilitate feasible implementations in experimental settings. They have also been explored as tools to devise schemes of quantum simulations of systems that are natively too difficult to tackle, say, quantum systems featuring certain kinds of intrinsic topological order [20,49]. At the same time, perturbative gadgets are quite often accompanied by unfavorable trade-offs in interaction strengths, recursions, or limitations to settings allowing for subspace restrictions, bringing into question their use for practical applications [11].

In this work, we have proposed a perturbative gadget that jointly lifts the requirements on subspace restrictions and recursions but cannot circumvent the problem of impractical interaction strengths. It achieves the same locality reduction as the subdivision gadget from Oliveira and Terhal [4] but has a direct formulation in higher-orders of perturbation theory as does the gadget from Jordan and Farhi [5], without requiring to restrict the evolution of the system to a subspace of the Hilbert space. In doing so, we observe that each of these constructions is a special case and demonstrate how to generalize our construction to a larger class of gadgets applicable to gate-based quantum computing.

Furthermore, we turn towards the setting of variational quantum algorithms, where the locality of cost functions can be of great interest and thus benefit from generic procedures producing Hamiltonians with the same ground states but reduced locality. Also for fault-tolerant quantum algorithms, reductions in locality could help alleviate hardware constraints or impact measurement noise.

We note that locality versus energy scale trade-offs in using perturbative gadgets may still be impractical outside of theoretical considerations and therefore leave the question of their practicality wide open. Circumventing the need for subspace restrictions and recursions does, however, broaden their scope and therefore opens up new directions of exploration for finding practical applications. It might be interesting to consider the question of how imposing restrictions on the target Hamiltonian might allow us to improve the resource requirements of our gadget construction. In the same flavor as gadgets for topological quantum codes, there might be classes of Hamiltonians for which the effective Hamiltonian can be retrieved at a lower perturbation order. Such a development would trade off universality for performance, an idea established in the area of variational quantum algorithms and beyond.

The numerical simulations were performed using the cross-platform Python library for differentiable programming of quantum computers PennyLane [50] and the code is available online [51].

been provided by the BMWK (PlanQK, EniQmA), the QuantERA (HQCC), the BMBF (Hybrid), the Munich Quantum Valley (K-8), the ERC (DebuQC), and the Einstein Foundation (Einstein Research Unit on Quantum Devices).

APPENDIX A: REVIEW OF PERTURBATION THEORY

The appendices are organized as follows: We begin reviewing the relevant results from perturbation theory in Appendix A. Then, we present the proof of Theorem 1 in Appendix B, discussing useful properties, the case of a simplified target Hamiltonian as well as the general case, before proving Corollary 1 in Appendix C. Appendix D focuses on extending the locality gadget to settings with mixed Pauli weights and for the reduction to an arbitrary target locality $k' > 3$. In Appendix E, we discuss whether perturbative gadgets can address the barren plateau problem in variational algorithms, while Appendix F outlines why the original gadget from Jordan and Farhi (Ref. [5]) is inapplicable in this setting. Numerical simulations are presented and discussed in Appendix G. Finally, we conclude by presenting another gadget construction inspired by the recipe, aimed at reducing the number of required measurement bases in the Hamiltonian readout problem in Appendix H.

In Appendix A, we briefly summarize relevant results from perturbation theory. In particular, we closely follow the presentation in Ref. [22] for obtaining a general perturbative expansion, which we need for the proof of Theorem 1 in Appendix B.

1. Definitions

The goal of perturbation theory is to calculate the effect of a small perturbation V of strength $\lambda \geqslant 0$ on a known, unperturbed Hamiltonian $H^{(0)}$. In particular, the goal is to study the Hamiltonian

$$H = H^{(0)} + \lambda V \tag{A1}$$

and to determine its low-energy eigenvalues and corresponding eigenvectors as a function of λ. To this end, suppose that $H^{(0)}$ and V act on a D-dimensional Hilbert space. Let $E_0^{(0)}$ be the lowest eigenenergy of $H^{(0)}$, and let

$$\mathcal{E}^{(0)} := \operatorname{span}\left\{ |\varphi\rangle \in \mathbb{C}^D : H^{(0)}|\varphi\rangle = E_0^{(0)}|\varphi\rangle \right\} \tag{A2}$$

be the corresponding d-dimensional (degenerate) ground space of $H^{(0)}$, $d \leqslant D$. We let $\{|j\rangle \in \mathbb{C}^D\}_{j=1}^d$ denote an orthonormal basis for $\mathcal{E}^{(0)}$. We also let $|\psi_1\rangle, |\psi_2\rangle, \ldots, |\psi_d\rangle \in \mathbb{C}^D$ be the perturbed orthonormal eigenvector (i.e., the eigenvectors of H) corresponding to the d-dimensional subspace $\mathcal{E}^{(0)}$, such that

$$H|\psi_j\rangle = \widetilde{E}_j|\psi_j\rangle \ \forall \ j \in \{1, 2, \ldots, d\}, \tag{A3}$$

where $\widetilde{E}_j$ are the energies corresponding to $|\psi_j\rangle$. For the rest of this section, we are concerned with the shifted energies

$E_j := \widetilde{E}_j - E_0^{(0)}$. We also let

$$\mathcal{E} := \operatorname{span}\{|\psi_j\rangle \in \mathbb{C}^D : j \in \{1, 2, \ldots, d\}\} \tag{A4}$$

be the d-dimensional subspace spanned by the vectors $|\psi_1\rangle, |\psi_2\rangle, \ldots, |\psi_d\rangle$. We assume throughout that the subspaces $\mathcal{E}^{(0)}$ and $\mathcal{E}$ are not orthogonal, meaning that there does not exist a vector in $\mathcal{E}^{(0)}$ that is orthogonal to all vectors in $\mathcal{E}$, and vice versa. This condition can be guaranteed if the perturbation strength is not too high [22].

Our specific goal in this section is to determine a perturbative expansion of H in the d-dimensional subspace $\mathcal{E}$, i.e., we seek a perturbative expansion of the *effective Hamiltonian*

$$H_{\text{eff}}(H, d) := \sum_{j=1}^d E_j|\psi_j\rangle\langle\psi_j|. \tag{A5}$$

Note that this is nothing more than the projection of the Hamiltonian H onto the subspace $\mathcal{E}$, followed by a constant shift of $E_0^{(0)}$.

Let P_0 be the projector onto $\mathcal{E}^{(0)}$. We can thus write $H^{(0)}$ as

$$H^{(0)} = E_0^{(0)}P_0 + Q, \tag{A6}$$

where Q satisfies $P_0 Q = Q P_0 = 0$. Let us also define $|\alpha_j\rangle$ to be the (non-normalized) state vectors reflecting the projections of the perturbed eigenstates onto the subspace $\mathcal{E}^{(0)}$, i.e.,

$$|\alpha_j\rangle := P_0|\psi_j\rangle. \tag{A7}$$

Note that these vectors are not necessarily orthonormal. But they are linearly independent, as we now prove.

Lemma 1 (linear independence). Under the assumption that the vector spaces $\mathcal{E}^{(0)}$ and $\mathcal{E}$ are not orthogonal, the vectors defined in Eq. (A7) are linearly independent. Consequently, there exist vectors $|\widetilde{\alpha}_j\rangle$, $j \in \{1, 2, \ldots, d\}$, such that $\langle\alpha_j|\widetilde{\alpha}_{j'}\rangle = \delta_{j,j'}$ for all $j, j' \in \{1, 2, \ldots, d\}$ and

$$P_0 = \sum_{j=1}^d |\alpha_j\rangle\langle\widetilde{\alpha}_j| = \sum_{j=1}^d |\widetilde{\alpha}_j\rangle\langle\alpha_j|. \tag{A8}$$

Proof. Consider the equation $c_1|\alpha_1\rangle + c_2|\alpha_2\rangle + \cdots + c_d|\alpha_d\rangle = 0$, where $c_1, c_2, \ldots, c_d \in \mathbb{C}$. Using the definition of $|\alpha_j\rangle$, we obtain

$$P_0(c_1|\psi_1\rangle + c_2|\psi_2\rangle + \cdots + c_d|\psi_d\rangle) = 0. \tag{A9}$$

Now, because $\mathcal{E}^{(0)}$ and $\mathcal{E}$ are nonorthogonal, it follows that the vector $c_1|\psi_1\rangle + c_2|\psi_2\rangle + \cdots + c_d|\psi_d\rangle$ is not in the orthogonal complement of $\mathcal{E}^{(0)}$, which means that it must be equal to the zero vector. Consequently, because the vectors $|\psi_j\rangle$ are linearly independent, we have that $c_1 = c_2 = \cdots = c_d = 0$. Hence, because the numbers $c_1, c_2, \ldots, c_d$ were arbitrary, we conclude that the vectors $|\alpha_j\rangle$ are linearly independent.

Now, consider the operator $T := \sum_{j=1}^d |\alpha_j\rangle\langle j|$, which can be thought of as a matrix whose columns are equal to the vectors $|\alpha_j\rangle$. In particular, note that $|\alpha_j\rangle = T|j\rangle$ for all $j \in \{1, 2, \ldots, d\}$. Linear independence of the vectors $|\alpha_j\rangle$ implies that T is invertible. (Here, and throughout this section, by *inverse* we mean the Moore–Penrose pseudo-inverse, which is the inverse on the support of a linear operator.) Hence, letting

$$|\widetilde{\alpha}_j\rangle := (T^{-1})^\dagger|j\rangle \ \forall \ j \in \{1, 2, \ldots, d\}, \tag{A10}$$

we find that

$$
\begin{aligned}
\langle \alpha_j | \widetilde{\alpha}_{j'} \rangle &= \langle j | T^\dagger (T^{-1})^\dagger | j' \rangle \\
&= \langle j | j' \rangle = \delta_{j,j'} \; \forall \, j, j' \in \{1, 2, \dots, d\},
\end{aligned} \quad \text{(A11)}
$$

as required.

Finally, using Eq. (A10) and the fact that $T|j\rangle = |\alpha_j\rangle$ for all $j \in \{1, 2, \dots, d\}$, we find that both $\sum_{j=1}^d |\alpha_j\rangle\langle\widetilde{\alpha}_j|$ and $\sum_{j=1}^d |\widetilde{\alpha}_j\rangle\langle\alpha_j|$ are equal to the identity operator for the subspace $\mathcal{E}^{(0)}$, which is precisely equal to the projection P_0. This completes the proof. ∎

Lemma 2 (invertibility). Consider the two linear operators

$$
\mathcal{U} := \sum_{j=1}^d |\psi_j\rangle\langle\widetilde{\alpha}_j|, \quad \text{(A12)}
$$

$$
\mathcal{A} := \lambda P_0 V \mathcal{U}. \quad \text{(A13)}
$$

The operator $\mathcal{U}$ is invertible on the subspace $\mathcal{E}$, and

$$
\mathcal{U} \mathcal{A} \mathcal{U}^{-1} = H_{\text{eff}}(H, d) = \sum_{j=1}^d E_j |\psi_j\rangle\langle\psi_j|. \quad \text{(A14)}
$$

Proof. First of all, note that

$$
\mathcal{U}|\alpha_j\rangle = |\psi_j\rangle \; \forall \, j \in \{1, 2, \dots, d\}, \quad \text{(A15)}
$$

which holds due to Lemma A 1. Next, using Eq. (A10), we find that $\mathcal{U} = \sum_{j=1}^d |\psi_j\rangle\langle j| T^{-1} = ST^{-1}$, where $S := \sum_{j=1}^d |\psi_j\rangle\langle j|$. Since the vectors $|\psi_j\rangle$ are linearly independent, the operator S is invertible. Furthermore, because the vectors $|\psi_j\rangle$ are orthonormal, $S^{-1} = S^\dagger = \sum_{j=1}^d |j\rangle\langle\psi_j|$. Therefore, the inverse of $\mathcal{U}$ exists and is equal to $\mathcal{U}^{-1} = TS^\dagger$. Next, using Eq. (A6), we have that

$$
P_0(\lambda V) = P_0\big(H - E_0^{(0)} P_0 - Q\big) = P_0 H - E_0^{(0)} P_0, \quad \text{(A16)}
$$

so that

$$
\mathcal{A} = P_0 H \mathcal{U} - E_0^{(0)} P_0 \mathcal{U}. \quad \text{(A17)}
$$

However, using Eq. (A8), we find that $P_0 \mathcal{U} = P_0$, which means that

$$
\mathcal{A} = P_0 H \mathcal{U} - E_0^{(0)} P_0. \quad \text{(A18)}
$$

For this reason, because $\mathcal{U}\mathcal{U}^{-1}$ is equal to the projection onto the subspace $\mathcal{E}$, we obtain

$$
\begin{aligned}
\mathcal{U} \mathcal{A} \mathcal{U}^{-1} &= \mathcal{U} P_0 H \mathcal{U} \mathcal{U}^{-1} - E_0^{(0)} \mathcal{U} P_0 \mathcal{U}^{-1} \\
&= \mathcal{U} P_0 \sum_{j=1}^d \widetilde{E}_j |\psi_j\rangle\langle\psi_j| - E_0^{(0)} \mathcal{U} P_0 \mathcal{U}^{-1} \\
&= \sum_{j=1}^d \widetilde{E}_j |\psi_j\rangle\langle\psi_j| - E_0^{(0)} \mathcal{U} P_0 \mathcal{U}^{-1}.
\end{aligned} \quad \text{(A19)}
$$

Finally, using Eq. (A8), along with Eq. (A15) and the fact that $\langle\widetilde{\alpha}_j| \mathcal{U}^{-1} = \langle j| T^{-1} T S^\dagger = \langle\psi_j|$, we find that

$$
\mathcal{U} P_0 \mathcal{U}^{-1} = \sum_{j=1}^d \mathcal{U}|\alpha_j\rangle\langle\widetilde{\alpha}_j| \mathcal{U}^{-1} = \sum_{j=1}^d |\psi_j\rangle\langle\psi_j|. \quad \text{(A20)}
$$

Therefore,

$$
\mathcal{U} \mathcal{A} \mathcal{U}^{-1} = \sum_{j=1}^d \big(\widetilde{E}_j - E_0^{(0)}\big)|\psi_j\rangle\langle\psi_j| = \sum_{j=1}^d E_j |\psi_j\rangle\langle\psi_j|, \quad \text{(A21)}
$$

as required. ∎

2. Perturbative expansion of $\mathcal{U}$

We now derive a perturbative expansion for $\mathcal{U}$, from which we obtain an expansion for $\mathcal{A}$, which will allow us to obtain a perturbative expansion for the effective Hamiltonian $H_{\text{eff}}(H, d)$ via Eq. (A14). Crucial to obtaining the perturbative expansion of $\mathcal{U}$ is the following fact:

Lemma 3 (towards a perturbative expansion). The linear operator $\mathcal{U}$ defined in Eq. (A12) satisfies

$$
\mathcal{U} = P_0 + A^{-1} Q_0 \lambda (V \mathcal{U} - \mathcal{U} V \mathcal{U}), \quad \text{(A22)}
$$

where $Q_0 := \mathbb{1} - P_0$ and $A := E_0^{(0)} \mathbb{1} - H^{(0)}$.

Proof. From Eq. (A3), we have that

$$
\big(H - E_0^{(0)} \mathbb{1}\big)|\psi_j\rangle = E_j |\psi_j\rangle \; \forall \, j \in \{1, 2, \dots, d\}. \quad \text{(A23)}
$$

Multiplying both sides of this equation from the left by P_0 gives us

$$
P_0\big(H - E_0^{(0)} \mathbb{1}\big)|\psi_j\rangle = P_0 E_j |\psi_j\rangle, \quad \text{(A24)}
$$

$$
\Rightarrow P_0\big(H^{(0)} + \lambda V - E_0^{(0)} \mathbb{1}\big)|\psi_j\rangle = E_j |\alpha_j\rangle, \quad \text{(A25)}
$$

$$
\Rightarrow \lambda P_0 V |\psi_j\rangle = E_j |\alpha_j\rangle, \quad \text{(A26)}
$$

for all $j \in \{1, 2, \dots, d\}$, where to obtain the right-hand side of the second line we used Eq. (A7). Now, multiplying both sides of the last line by $\mathcal{U}$ gives us

$$
\lambda \mathcal{U} P_0 V |\psi_j\rangle = E_j \mathcal{U}|\alpha_j\rangle, \quad \text{(A27)}
$$

$$
\Rightarrow \lambda \mathcal{U} V |\psi_j\rangle = E_j |\psi_j\rangle, \quad \text{(A28)}
$$

for all $j \in \{1, 2, \dots, d\}$, where to obtain the left-hand side of the last line we used the fact that $\mathcal{U} P_0 = \mathcal{U}$, which can be straightforwardly verified using Eq. (A8). For the right-hand side of the last line, we used Eq. (A15) the fact that $\mathcal{U}|\alpha_j\rangle = |\psi_j\rangle$ for all $j \in \{1, 2, \dots, d\}$. Using Eq. (A23), we find that Eq. (A28) can be written as

$$
\lambda \mathcal{U} V |\psi_j\rangle = \big(H^{(0)} - E_0^{(0)} \mathbb{1}\big)|\psi_j\rangle, \quad \text{(A29)}
$$

$$
\Rightarrow \big(H - E_0^{(0)} \mathbb{1} - \lambda \mathcal{U} V\big)|\psi_j\rangle = 0, \quad \text{(A30)}
$$

for all $j \in \{1, 2, \dots, d\}$. Multiplying the last line by $\langle\widetilde{\alpha}_j|$ from the right leads to

$$
\big(H - E_0^{(0)} \mathbb{1} - \lambda \mathcal{U} V\big)|\psi_j\rangle\langle\widetilde{\alpha}_j| = 0 \; \forall \, j \in \{1, 2, \dots, d\}. \quad \text{(A31)}
$$

Summing over all $j \in \{1, 2, \dots, d\}$ therefore leads to

$$
\big(H - E_0^{(0)} \mathbb{1} - \lambda \mathcal{U} V\big)\mathcal{U} = 0, \quad \text{(A32)}
$$

which is equivalent to

$$
\big(E_0^{(0)} \mathbb{1} - H^{(0)}\big)\mathcal{U} = \lambda V \mathcal{U} - \lambda \mathcal{U} V \mathcal{U}. \quad \text{(A33)}
$$

Letting $Q_0 := \mathbb{1} - P_0$ be the projector onto the span of the excited states of $H^{(0)}$, we have

$$
\mathcal{U} = P_0 \mathcal{U} + Q_0 \mathcal{U}, \quad \text{(A34)}
$$

which means that

$$\left(E_0^{(0)}\mathbb{1} - H^{(0)}\right)\mathcal{U} = \left(E_0^{(0)}\mathbb{1} - H^{(0)}\right)(P_0\mathcal{U} + Q_0\mathcal{U}$$
$$= 0 + \left(E_0^{(0)}\mathbb{1} - H^{(0)}\right)Q_0\mathcal{U}, \quad \text{(A35)}$$

leading to

$$\left(E_0^{(0)}\mathbb{1} - H^{(0)}\right)Q_0\mathcal{U} = \lambda V\mathcal{U} - \lambda \mathcal{U}V\mathcal{U}. \quad \text{(A36)}$$

Next, using the fact that $P_0\mathcal{U} = P_0$, we obtain

$$Q_0(V\mathcal{U} - \mathcal{U}V\mathcal{U}) = (\mathbb{1} - P_0)(V\mathcal{U} - \mathcal{U}V\mathcal{U})$$
$$= V\mathcal{U} - \mathcal{U}V\mathcal{U} + P_0(\mathcal{U}V\mathcal{U} - V\mathcal{U})$$
$$= V\mathcal{U} - \mathcal{U}V\mathcal{U}. \quad \text{(A37)}$$

Then, because $H^{(0)} - E_0^{(0)}\mathbb{1}$ has a well-defined inverse on $\mathcal{E}^{(0)\perp}$, we can write

$$Q_0\mathcal{U} = \left(E_0^{(0)}\mathbb{1} - H^{(0)}\right)^{-1}Q_0\lambda(V\mathcal{U} - \mathcal{U}V\mathcal{U}). \quad \text{(A38)}$$

Inserting this result into Eq. (A34) yields

$$\mathcal{U} = P_0\mathcal{U} + \left(E_0^{(0)}\mathbb{1} - H^{(0)}\right)^{-1}Q_0\lambda(V\mathcal{U} - \mathcal{U}V\mathcal{U}), \quad \text{(A39)}$$

which simplifies to

$$\mathcal{U} = P_0 + \left(E_0^{(0)}\mathbb{1} - H^{(0)}\right)^{-1}Q_0\lambda(V\mathcal{U} - \mathcal{U}V\mathcal{U}), \quad \text{(A40)}$$

as required. ∎

We have, therefore, obtained the governing equation for $\mathcal{U}$. This equation is well suited for expansion in powers of λ and can be expanded as

$$\mathcal{U} = \sum_{m=0}^{\infty} \mathcal{U}^{(m)}, \quad \text{(A41)}$$

with $\mathcal{U}^{(m)}$ being the mth-order term. Substituting $\mathcal{U}$ into Eq. (A22) gives the recurrence relations

$$\mathcal{U}^{(0)} = P_0, \quad \text{(A42)}$$

$$\mathcal{U}^{(m)} = A^{-1}Q_0\lambda\left(V\mathcal{U}^{(m-1)} - \sum_{p=1}^{m-1}\mathcal{U}^{(p)}V\mathcal{U}^{(m-p-1)}\right), \quad \text{(A43)}$$

with $A = E_0^{(0)}\mathbb{1} - H^{(0)}$. Let

$$S^\ell := \begin{cases} -P_0 & \text{if } \ell = 0 \\ A^{-\ell}Q_0 & \text{if } \ell > 0. \end{cases} \quad \text{(A44)}$$

Then,

$$\mathcal{U}^{(m)} = \lambda^m {\sum}' S^{\ell_1}VS^{\ell_2}V\cdots VS^{\ell_n}VP_0, \quad \text{(A45)}$$

where the sum is over all sets of indices $\{\ell_i\}_{i=1}^m$ that fulfill the following conditions [22]:

$$\ell_i \geqslant 0 \,\forall\, i \in \{1,\dots,m\}, \quad \text{(A46)}$$

$$\ell_1 + \cdots + \ell_{m-1} = m, \quad \text{(A47)}$$

$$\ell_1 + \cdots + \ell_p \geqslant p \,\forall\, p \in \{1,\dots,m-1\}. \quad \text{(A48)}$$

These conditions can be understood as all the possible paths on a square lattice of length m that go from the bottom left corner to the top right one and lie above the diagonal, as represented in Fig. 4.
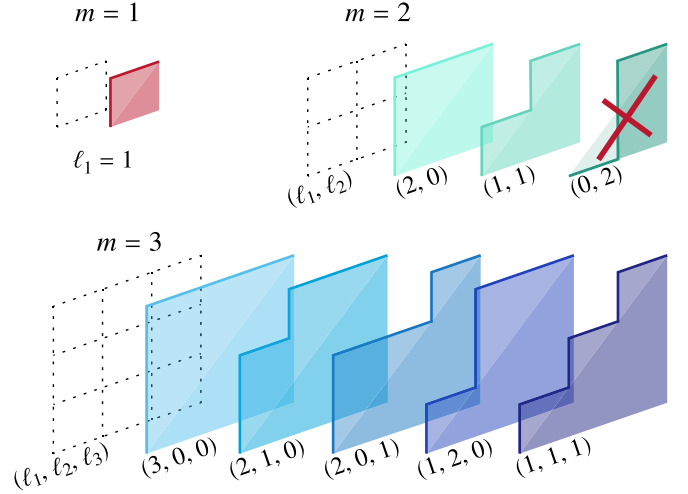

FIG. 4. Visual representation of the allowed set of indices in the expansion of $\mathcal{U}$ as proposed by Bloch in Ref. [22]. All allowed diagrams are shown for $m \in \{1, 2, 3\}$ and additionally, one not contributing diagram for $m = 2$ since it goes below the diagonal [it does not fulfill the third property in Eq. (A46)].

Lemma 4 (explicit form of terms). For $\ell > 0$, the terms S^ℓ defined in Eq. (A44) are given by

$$S^\ell = \sum_{j \neq 0} \left(E_0^{(0)} - E_j^{(0)}\right)^{-\ell}P_j \quad (\ell > 0), \quad \text{(A49)}$$

where P_j is the spectral projection of $H^{(0)}$ corresponding to the energy $E_j^{(0)}$.

Proof. To obtain this result, we use the fact that

$$H^{(0)} = E_0^{(0)}P_0 + \sum_{j \neq 0} E_j^{(0)}P_j, \quad P_0 + \sum_{j \neq 0} P_j = \mathbb{1}. \quad \text{(A50)}$$

This implies that

$$E_0^{(0)}\mathbb{1} - H^{(0)} = \sum_{j \neq 0}\left(E_0^{(0)} - E_j^{(0)}\right)P_j, \quad \text{(A51)}$$

which in turn implies that

$$A^{-\ell} = \left(E_0^{(0)}\mathbb{1} - H^{(0)}\right)^{-\ell} = \sum_{j \neq 0}\left(E_0^{(0)} - E_j^{(0)}\right)^{-\ell}P_j, \quad \text{(A52)}$$

and because this is supported entirely on the orthogonal complement of the ground space of $H^{(0)}$, we have that $A^{-\ell}Q_0 = A^{-\ell}$, proving the desired result. ∎

From the expansion of $\mathcal{U}$, we can write the expansion of $\mathcal{A}$ as

$$\mathcal{A} = \sum_{m=1}^{\infty} \mathcal{A}^{(m)}, \quad \text{(A53)}$$

with

$$\mathcal{A}^{(m)} = \lambda^m {\sum}'' P_0 V S^{\ell_1} V S^{\ell_2} V \cdots V S^{\ell_{m-1}} V P_0, \quad \text{(A54)}$$

with the sum over all sets of $m - 1$ indices $\{\ell_i\}_{i=1}^{m-1}$ such that

$$\ell_i \geqslant 0 \,\forall\, i \in \{1, 2, \dots, m-1\}, \quad \text{(A55)}$$

$$\ell_1 + \cdots + \ell_{m-1} = m - 1 \geqslant 0, \quad \text{(A56)}$$

$$\ell_1 + \cdots + \ell_p \geqslant p \,\forall\, p \in \{1, \dots, m-2\}. \quad \text{(A57)}$$

Finally, using Eq. (A14), the perturbative expansion of $H_{\mathrm{eff}}(H, d)$, up to mth order in perturbation theory, is

$$H_{\mathrm{eff}}(H, d) = \mathcal{U}\mathcal{A}^{(\leqslant m)}\mathcal{U}^{-1} + O(\lambda^{m+1}) \quad \mathcal{A}^{(\leqslant m)} := \sum_{j=1}^{m} \mathcal{A}^{(j)}. \tag{A58}$$

As shown in Ref. [22] (see also Ref. [5, Appendix B]), the following condition is sufficient to guarantee convergence of this perturbative expansion:

$$\|\lambda V\| < \frac{\gamma}{4}, \tag{A59}$$

where γ is the gap between $E_0^{(0)}$ and the energy of the first-excited state of $H^{(0)}$. We note that this bound is not tight [17] and that significantly larger λ might still lead to similar results, even outside of the regime in which the perturbative expansion converges [14,19].

APPENDIX B: PROOF OF THEOREM 1

In this section, we show how to get from Definition 1 to Theorem 1 using the expansion presented in the previous section. In particular, we analyze what the formulas obtained in Appendix A imply for the construction of the perturbative gadget introduced in Definition 1, and thereby prove Theorem 1. The proof also serves as the groundwork for the extensions presented in Appendix D and the generalization in Sec. IV.

For convenience, let us recall the definition of our gadget Hamiltonian from Definition 1. Given a k-body target Hamiltonian H^{target} acting on a Hilbert space $\mathcal{H}^{\mathrm{target}}$ with $\dim(\mathcal{H}^{\mathrm{target}}) = 2^n$, the associated gadget Hamiltonian is

$$H^{\mathrm{gad}} := \sum_{s=1}^{r} H_s^{\mathrm{aux}} + \lambda \sum_{s=1}^{r} V_s, \tag{B1}$$

where

$$H_s^{\mathrm{aux}} := \sum_{j=1}^{k} \frac{1}{2}\left(\mathbb{1}_{s,j} - Z_{s,j}^{\mathrm{aux}}\right) = \sum_{j=1}^{k} |1\rangle\langle 1|_{s,j}^{\mathrm{aux}}, \tag{B2}$$

$$V_s := \sum_{j=1}^{k} \tilde{c}_{s,j} \sigma_{s,j}^{\mathrm{target}} \otimes X_{s,j}^{\mathrm{aux}} \otimes X_{s,(j+1) \bmod k}^{\mathrm{aux}}, \tag{B3}$$

and $\tilde{c}_{s,j} = -(-1)^k c_s$ if $j = 1$ or $\tilde{c}_{s,j} = 1$ otherwise. We consider $\sigma_{s,j} \in \{\hat{n} \cdot \boldsymbol{\sigma} : \hat{n} \in \mathbb{R}^3, \hat{n} \cdot \hat{n} = 1, \boldsymbol{\sigma} = (X, Y, Z)\}$. We refer to the Hilbert spaces of each of the auxiliary registers as $\mathcal{H}_s^{\mathrm{aux}}$, where $\dim(\mathcal{H}_s^{\mathrm{aux}}) = 2^k$. The complete Hilbert space acted upon by our gadget Hamiltonian is thus

$$\mathcal{H}^{\mathrm{total}} \simeq \mathcal{H}^{\mathrm{target}} \otimes \mathcal{H}_1^{\mathrm{aux}} \otimes \cdots \otimes \mathcal{H}_r^{\mathrm{aux}}. \tag{B4}$$

For an illustration of the different registers in the case of a toy example, see Fig. 5. Our goal in this section is to use the perturbative expansion of the previous section, by taking $H^{(0)} \equiv \sum_{s=1}^{r} H_s^{\mathrm{aux}}$ and $V \equiv \sum_{s=1}^{r} V_s$, and showing that

$$H_{\mathrm{eff}}(H^{\mathrm{gad}}, 2^n) = a H^{\mathrm{target}} \otimes |\phi_0\rangle\langle\phi_0| + b\Pi + O_{\mathrm{err}}, \tag{B5}$$

where Π is the projector onto the low-energy subspace of H^{gad}, $|\phi_0\rangle \in \mathcal{H}^{\mathrm{aux}}$ is a state vector, a is a scaling factor, and b is a shift on the whole subspace of interest. We remark that the above equation is in accordance with [19, Definition 11], which provides a general definition of what it means for one
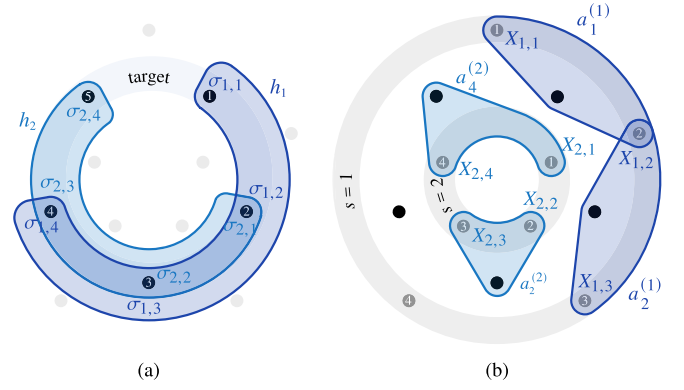

FIG. 5. Illustration of the different relevant qubit registers in the construction of the gadget Hamiltonian for the case of $n = 5$, $k = 4$, and $r = 2$. Shown in panel (a) are the two target Hamiltonian terms h_1 and h_2 on the five-qubit target register with the corresponding single-qubit operators $\sigma_{s,j}$. Panel (b) shows the two four-qubit auxiliary registers, and two of the $a_j^{(s)}$ operators displayed on each. As displayed for $\sigma_{2,3}$, the indices (s, j) for operators on the target register should always be understood by referring to h_s. Indeed, $\sigma_{2,3}$ acts on qubit 4. On the other hand, the indices for operators on the auxiliary qubits are more straightforward: $X_{2,3}$ is a Pauli X on the third qubit of the second register.

Hamiltonian to mimic another Hamiltonian in its low-energy subspace.

1. Useful properties

For illustrative purposes, and to help in some coming steps of the derivation, let us write down the lowest-order terms of the expansion of $\mathcal{A}$ from Eq. (A54) explicitly. This is

$$\mathcal{A}^{(1)} = \lambda P_0 V P_0, \tag{B6}$$

$$\mathcal{A}^{(2)} = \lambda^2 P_0 V S^1 V P_0, \tag{B7}$$

$$\mathcal{A}^{(3)} = \lambda^3 P_0 V S^1 V S^1 V P_0 + \lambda^3 P_0 V S^2 V S^0 V P_0. \tag{B8}$$

Let us consider a few properties of the construction defined in Definition 1, which we use below. First of all, note that the gadget Hamiltonian is three-body by construction. Then, the unperturbed Hamiltonian acting solely on the auxiliary registers can be rewritten as

$$H^{\mathrm{aux}} = \sum_{s=1}^{r} \sum_{j=1}^{k} |1\rangle\langle 1|_{s,j} \tag{B9}$$

and its ground state is the computational basis state vector $|0\rangle^{\otimes rk}$ with corresponding eigenvalue 0. Consequently, the ground space of the unperturbed Hamiltonian is given by

$$\mathcal{E}^{(0)} = \mathrm{span}\{|\varphi\rangle \otimes |0\rangle^{\otimes rk} : |\varphi\rangle \in \mathcal{H}^{\mathrm{target}}\}, \tag{B10}$$

with the corresponding projector

$$P_0 = \mathbb{1}^{\otimes n} \otimes (|0\rangle\langle 0|)^{\otimes rk}, \tag{B11}$$

whose support is 2^n-dimensional, as for the original target Hamiltonian. Furthermore, the energy of any state in the

computational basis is given by the Hamming weight of the state with respect to the auxiliary register. Lastly, let us define

$$a_j^{(s)} := X_{s,j}^{\text{aux}} \otimes X_{s,(j+1) \bmod k}^{\text{aux}} \tag{B12}$$

as the parts of the perturbation from Eq. (B3) acting on the auxiliary registers. Then, these operators fulfill the useful relations

$$\left[a_i^{(s)}, a_j^{(p)}\right] = 0 \;\forall\, i, j, s, p, \quad \prod_{j=1}^{k} a_j^{(s)} = \bigotimes_{j=1}^{k} X_{s,j}^2 = \mathbb{1}^{\otimes k} \;\forall\, s. \tag{B13}$$

The first is a direct consequence of having only Pauli-X operators in the construction of $a_j^{(s)}$. The second is due to the fact that the operators $a_j^{(s)}$ are constructed in a cyclic manner and the fact that the Pauli-X operator squares to the identity.

2. Simplified target Hamiltonian

Now, let us first consider the simplified example of a target Hamiltonian comprising only a single term with unit norm, that is

$$H^{\text{target}} = \sigma_1 \sigma_2 \cdots \sigma_k. \tag{B14}$$

For this case, we can omit the subscript s. The corresponding gadget Hamiltonian is then given by

$$H^{\text{gad}} = H^{\text{aux}} + \lambda V, \tag{B15}$$

with

$$H^{\text{aux}} = \sum_{j=1}^{k} \frac{1}{2} (\mathbb{1} - Z_j), \tag{B16}$$

and

$$V = \sum_{j=1}^{k} (-(-1)^k)^{\delta_{1j}} \sigma_j \otimes X_j \otimes X_{(j+1) \bmod k}. \tag{B17}$$

The 2^k-dimensional unperturbed ground space is

$$\mathcal{E}^{(0)} = \text{span}\{|\varphi\rangle^{\text{target}} \otimes |0\rangle^{\otimes k} : |\varphi\rangle \in \mathcal{H}^{\text{target}}\}. \tag{B18}$$

The projector on the unperturbed ground space can consequently be written as

$$P_0 = \mathbb{1}^{\otimes k} \otimes (|0\rangle\langle 0|)^{\otimes k}. \tag{B19}$$

First, let us look at the auxiliary part of the Hamiltonian to understand its effect. It can be interpreted as a penalization on flipped qubits: each term has a zero contribution if the affected qubit is in the $|0\rangle$ state but has a penalty of one if the qubit is in the $|1\rangle$ state. The gap is then $\gamma = 1$ and taking into account that $\|V\| = k$, the convergence of the expansion in Eq. (A45) is guaranteed for $\lambda \leqslant 1/(4k)$.

Let us now study the first terms in the expansion of $\mathcal{A}$ from Eq. (A54), i.e.,

$$\mathcal{A}^{(\leqslant 2)} = \lambda P_0 V P_0 + \lambda^2 P_0 V S^1 V P_0. \tag{B20}$$

Every term of this expansion is sandwiched between two projectors P_0. The only terms with a nontrivial contribution are then those that take a state from $\mathcal{E}^{(0)}$ and return it to $\mathcal{E}^{(0)}$, i.e., those that leave the auxiliary register in the all-zero state. That is not the case for $\mathcal{A}^{(1)}$ in which the application of a

single term of the perturbation V necessarily kicks the state out of the ground space by flipping two auxiliary qubits. Consequently, $\mathcal{A}^{(1)} = 0$. For $\mathcal{A}^{(2)}$, each term of the first V excites the system to a state with two flipped auxiliary qubits. The only contribution from S^1 is then due to the projector on the second excited subspace P_2 with corresponding eigenenergy $E_2^{(0)} = 2$. This gives the denominator in S^1, resulting in

$$\mathcal{A}^{(\leqslant 2)} = -\frac{\lambda^2}{2} P_0 V^2 P_0. \tag{B21}$$

Now, let us examine V^2. This twofold application of the perturbation V has to act on the same two qubits, flipping them back to their original state. All cross terms in

$$\begin{aligned}
V^2 &= \sum_{i,j=1}^{k} \sigma_i \sigma_j \otimes a_i a_j \\
&= \sum_{i,j=1}^{k} \sigma_i \sigma_j \otimes X_i X_{(i+1) \bmod k} X_j X_{(j+1) \bmod k},
\end{aligned} \tag{B22}$$

with $i \neq j$, do not contribute since some of the Pauli-X operators on auxiliary qubits survive and leave the state outside of $\mathcal{E}^{(0)}$. For visualization, we refer to Fig. 3(c). Additionally, for $i = j$, and even on the target register, the terms that do contribute are necessarily squared operators and thus only identities:

$$\sum_{j=1}^{k} \sigma_j^2 \otimes (X_j X_{(j+1) \bmod k})^2 = \sum_{j=1}^{k} \mathbb{1} = k\mathbb{1}. \tag{B23}$$

Altogether, we have that

$$\mathcal{A}^{(\leqslant 2)} = -\frac{\lambda^2 k}{2} P_0. \tag{B24}$$

Similarly, for the next orders in the perturbative expansion, each pair of flipped auxiliary qubits has to be flipped back by the same operator in order to result in a nonzero contribution. For this to be true, m is only allowed to take even values, and the perturbation is always applied as $P_0 V^m P_0 \propto P_0$. That means that all contributions below order k are proportional to P_0 and thus

$$\mathcal{A}^{(\leqslant k-1)} = \left(\sum_{m=0}^{k-1} \alpha_m \lambda^m\right) P_0. \tag{B25}$$

Here, α_m is a coefficient depending on the number of combinations of m excitations returning a state to $\mathcal{E}^{(0)}$, and the corresponding energy penalties picked up along the way, which is independent of λ. To strengthen the intuition on the construction of α_m we refer to Fig. 6, which displays a few examples of terms that do and do not contribute.

A different behavior can be first observed at kth order in perturbation theory. There, by Eq. (B13), we can construct terms acting k times on the auxiliary register as $\prod_{j=1}^{k} a_j = \mathbb{1}^{\otimes k}$. In other words, all qubits from the auxiliary register are flipped twice and thus returned to $\mathcal{E}^{(0)}$, while each qubit in the target register is acted upon only once by the corresponding σ_j. This results in all elements from H^{target} being applied to
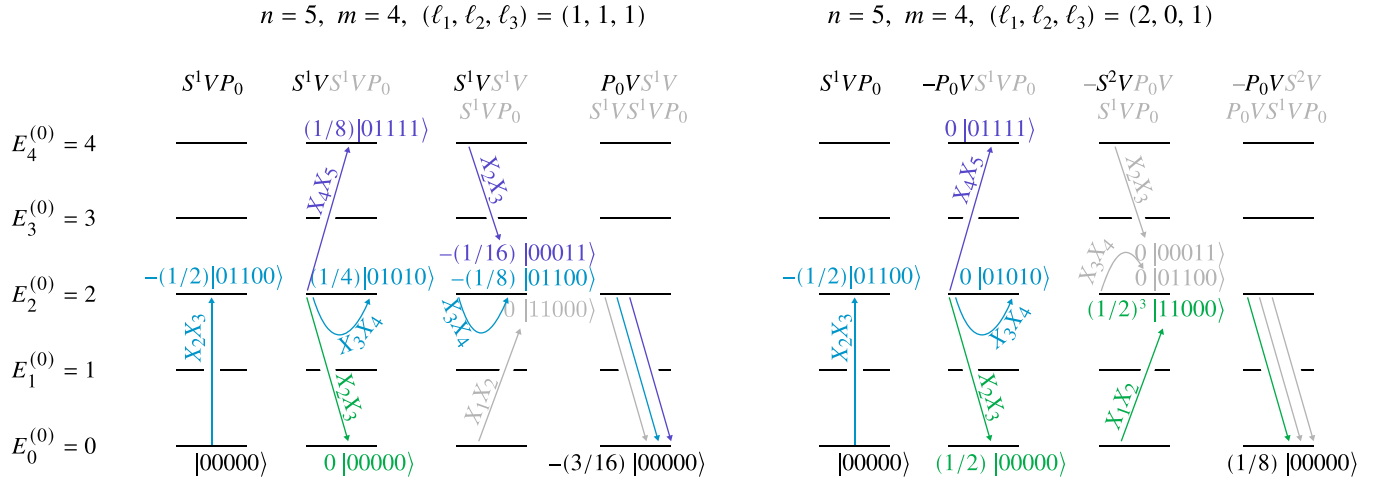
$$n = 5, \ m = 4, \ (\ell_1, \ell_2, \ell_3) = (1, 1, 1)$$

$$n = 5, \ m = 4, \ (\ell_1, \ell_2, \ell_3) = (2, 0, 1)$$

FIG. 6. Illustration of the interplay of V and S^ℓ in $\mathcal{A}$. Shown are some possible contributions to α_4 in the case of seven qubits; see Eq. (B25). Displayed are two choices of indices ℓ_i fulfilling (A55) and some terms from V^4 that starting from a state in $\mathcal{E}^{(0)}$ return to $\mathcal{E}^{(0)}$ for each choice. At each step of application of $S^{\ell_i}V$, the state is changed by V by flipping two qubits, and S^{ℓ_i} weights the resulting state by a power of the inverse of the energy or 0.

the target register, yielding

$$\mathcal{A}^{(k)} = -(-\lambda)^k \sum_{\boldsymbol{\ell}} \frac{1}{\xi_{\boldsymbol{\ell}}} P_0 (-(-1)^k \sigma_1 \sigma_2 \cdots \sigma_k \otimes \mathbb{1}^{\otimes k}) P_0$$

$$= \lambda^k \sum_{\boldsymbol{\ell}} \frac{1}{\xi_{\boldsymbol{\ell}}} P_0 (H^{\text{target}} \otimes \mathbb{1}^{\otimes k}) P_0. \tag{B26}$$

The sum over $\boldsymbol{\ell}$ is over all possible permutations in the application order of the (σ_j, a_j) pairs, while $\xi_{\boldsymbol{\ell}} = \prod_{i=1}^k E_{\ell_i}$ is the factor originating from the corresponding energy penalties $E_{\ell_i} = \|H^{\text{aux}} a_{\ell_i} \cdots a_{\ell_0} |0\rangle^{\otimes k}\|_2$. Defining

$$\Xi^{-1} := \sum_{\boldsymbol{\ell}} \frac{1}{\xi_{\boldsymbol{\ell}}}, \tag{B27}$$

the full expansion of $\mathcal{A}$ up to kth order in perturbation theory considering all possible combinations can then be written as

$$\mathcal{A}^{(\leqslant k)} = f(\lambda) P_0 + \frac{\lambda^k}{\Xi} P_0 (H^{\text{target}} \otimes \mathbb{1}^{\otimes k}) P_0$$

$$= f(\lambda) P_0 + \frac{\lambda^k}{\Xi} H^{\text{target}} \otimes (|0\rangle\langle 0|)^{\otimes k}. \tag{B28}$$

Applying these results to Eq. (A58) yields

$$H_{\text{eff}}(H^{\text{gad}}, 2^n) = \mathcal{U} f(\lambda) P_0 \mathcal{U}^{-1} + \mathcal{U} \left[\frac{\lambda^k}{\Xi} H^{\text{target}} \otimes (|0\rangle\langle 0|)^{\otimes k} \right.$$

$$\left. + O(\lambda^{k+1}) \right] \mathcal{U}^{-1}$$

$$= f(\lambda) \Pi + \frac{\lambda^k}{\Xi} H^{\text{target}} \otimes (|0\rangle\langle 0|)^{\otimes k} + O(\lambda^{k+1}), \tag{B29}$$

where $\mathcal{U} P_0 \mathcal{U}^{-1} = \Pi$ is the projector onto the support of $H_{\text{eff}}(H^{\text{gad}}, 2^n)$. Similarly to Ref. [5], we used the fact that the

operators $\mathcal{U}$ on both sides leave the second term unaffected up to errors of order $O(\lambda^{k+1})$.

3. General case

For the general case stated in Definition 1, the argument is similar, with the exception that one has to bear in mind that there are now r auxiliary registers of k qubits each. There are then more cross terms in the powers of V and Eq. (B22) becomes

$$V^2 = \sum_{s,p=1}^r \sum_{i,j=1}^k \tilde{c}_{s,i} c_{p,j} \sigma_{s,i} \sigma_{p,j} \otimes X_{s,i} X_{s,i+1} X_{p,j} X_{p,j+1}. \tag{B30}$$

Similarly, only terms with $s = p$ and $i = j$ contribute. More contributions appear from all possible combinations of the different terms in the powers of V, but in the end, Eq. (B25) still holds: All nonzero contributions to $\mathcal{A}$ at orders lower than k are proportional to P_0.

At kth order in perturbation theory, the nontrivial terms appear when acting k times on a single auxiliary register, since Eq. (B13) only holds when all operators operate on the same register s.

On the target register, it produces a contribution of the form $\prod_{j=1}^k \tilde{c}_{s,j} \sigma_{s,1} \sigma_{s,2} \cdots \sigma_{s,k} = -(-1)^k c_s h_s$. Considering all possible cross terms of this kind emerging from V^k leads to similar terms for each auxiliary register, yielding

$$\mathcal{A}^{(\leqslant k)} = f(\lambda) P_0 + \sum_{s=1}^r \frac{\lambda^k}{\Xi} P_0 (c_s h_s \otimes \mathbb{1}_s^{\otimes k}) P_0$$

$$= f(\lambda) P_0 + \frac{\lambda^k}{\Xi} P_0 \left(\sum_{s=1}^r c_s h_s \otimes \mathbb{1}_s^{\otimes k} \right) P_0$$

$$= f(\lambda) P_0 + \frac{\lambda^k}{\Xi} P_0 (H^{\text{target}} \otimes \mathbb{1}^{\otimes rk}) P_0. \tag{B31}$$

Applying these results to Eq. (A58) results in

$$
H_{\text{eff}}(H^{\text{gad}}, 2^n) = \mathcal{U}f(\lambda)P_0\mathcal{U}^{-1} + \mathcal{U}\left[\frac{\lambda^k}{\Xi}P_0(H^{\text{target}} \otimes \mathbb{1}^{\otimes rk})P_0\right.
$$
$$
\left. + O(\lambda^{k+1})\right]\mathcal{U}^{-1}
$$
$$
= f(\lambda)\Pi + \frac{\lambda^k}{\Xi}H^{\text{target}} \otimes (|0\rangle\langle 0|)^{\otimes rk} + O(\lambda^{k+1}),
$$

(B32)

which is what we have been aiming for. Furthermore, by Eq. (A59), we need to upper bound the perturbation strength λ by $\gamma/(4\|V\|)$. For the presented gadget we have $\gamma = 1$ and $\|V\| \leqslant [\sum_{s=1}^{r} |c_s| + r(k-1)]$ by the triangle inequality and the fact that the operator norm of Pauli operators is equal to one. Although not tight, we can thus upper bound λ by

$$
\lambda \leqslant \frac{1}{4[\sum_{s=1}^{r} |c_s| + r(k-1)]},
$$

(B33)

concluding the proof of Theorem 1. ∎

APPENDIX C: PROOF OF COROLLARY 1

Corollary 1 results from a direct application of Eq. (6) for sufficiently small λ. Since $f(\lambda)\Pi$ is a constant shift on the whole subspace of interest, the eigenstates of $H_{\text{eff}}(H^{\text{gad}}, 2^n)$ and $H_{\text{eff}}(H^{\text{gad}}, 2^n) - f(\lambda)\Pi$ are identical, and all energy gaps preserved. We can then rewrite Eq. (6) as

$$
\widetilde{H}_{\text{eff}}(H^{\text{gad}}, 2^n, f(\lambda)) = \frac{\lambda^k}{\Xi}H^{\text{target}} \otimes (|0\rangle\langle 0|)^{\otimes rk} + \lambda^{k+1}O^{\text{err}},
$$

(C1)

with $\widetilde{H}_{\text{eff}}(H^{\text{gad}}, 2^n, f(\lambda)) := H_{\text{eff}}(H^{\text{gad}}, 2^n) - f(\lambda)\Pi$ and $\|O^{\text{err}}\| \in O(1)$. Since the error shifts any eigenvalue by at most $\|\lambda^{k+1}O^{\text{err}}\|$, choosing λ such that

$$
\|\lambda^{k+1}O^{\text{err}}\| \leqslant \lambda^k\left(E_1^{\text{target}} - E_0^{\text{target}}\right)/\Xi,
$$

(C2)

where E_0^{target} and E_1^{target} are the ground and first-excited energies, respectively, of H^{target}, ensures that the ground space of the right-hand side of Eq. (C1) remains separated from the first excited subspace. Considering that the spectrum of H^{target} is independent of λ, we are always able to find a sufficiently small λ, whose upper bound is given by

$$
\lambda \leqslant \frac{E_1^{\text{target}} - E_0^{\text{target}}}{\Xi\|O^{\text{err}}\|},
$$

(C3)

concluding the proof. ∎

APPENDIX D: EXTENSIONS OF THE LOCALITY GADGET CONSTRUCTION

1. Extension to mixed Pauli weights

The gadget presented in Definition 1 seems to be restricted to the case in which all terms h_s of the target Hamiltonian act nontrivially on the same number of qubits, namely, that they all act nontrivially on exactly k qubits. Here, we want to argue that such a restriction, although useful to simplify the derivations and formulas, is not necessary. We can also consider the case of a Hamiltonian H^{target} whose terms h_s are

not all acting nontrivially on the same number of qubits, i.e.,

$$
H^{\text{target}} = \sum_{s=1}^{r} c_s h_s \quad \text{with} \quad h_s = \bigotimes_{j=1}^{k_s} \sigma_{s,j},
$$

(D1)

where k_s and $k_{s'}$ may differ. Let us, therefore, define $k = \max_s k_s$. This means that all terms h_s act nontrivially on at most k qubits.

First of all, one should note that no step of the derivation in Appendix B actually relies on the fact that $\sigma_{s,j}$ is a linear combination of the Pauli operators X, Y, Z. What has been leveraged, on the other hand, is that without loss of generality and to simplify the derivation, we can assume the property $\sigma_{s,j}^2 = \mathbb{1}$. In other words, nothing prohibits us from extending this definition to include identity terms, such that $\sigma_{s,j} \in \{\hat{\boldsymbol{n}} \cdot \boldsymbol{\sigma} : \hat{\boldsymbol{n}} \in \mathbb{R}^3, \hat{\boldsymbol{n}} \cdot \hat{\boldsymbol{n}} = 1, \boldsymbol{\sigma} = (X, Y, Z)\} \bigcup \{\mathbb{1}\}$. We can therefore redefine the gadget from Eq. (3) such that H_s^{aux} stays as in Definition 1, but the perturbation changes to

$$
V_s = \sum_{j=1}^{k_s} \tilde{c}_{s,j}\sigma_{s,j}^{\text{target}} \otimes X_{s,j}^{\text{aux}} \otimes X_{s,(j+1) \bmod k}^{\text{aux}}
$$
$$
+ \sum_{j=k_s+1}^{k} X_{s,j}^{\text{aux}} \otimes X_{s,(j+1) \bmod k}^{\text{aux}}.
$$

(D2)

With this change, even terms with $k_s < k$ will only contribute at kth order in perturbation theory, and we recover our main result even for Hamiltonians of the form presented in Eq. (D1).

2. Extension to arbitrary target locality

Having extended the use of our gadget to target Hamiltonians whose terms act nontrivially on different numbers of qubits, we can also extend it for arbitrary target locality. For now, let us assume again that all terms h_s act nontrivially on exactly k qubits. Reducing the locality to a value lower than three is out of reach for our gadget construction. However, we can construct gadget Hamiltonians with localities between three and k that lead to the same overall result as Theorem 1. Doing so might seem to be a counter-productive goal at first, but such constructions will require fewer additional qubits and might thus be useful when it comes to practical implementations. After all, the three-body perturbative gadget introduced here requires rk additional qubits. We start by discussing the simpler case in which k is divisible by $k' - 2$, where k' is the target locality, and then extend the findings to the general case.

If, for instance, one chooses the target locality to be four instead of three, Eq. (B3) can be changed to

$$
V_s = \sum_{j=1}^{k/2} \tilde{c}_{s,j}\sigma_{s,2j-1}^{\text{target}} \otimes \sigma_{s,2j}^{\text{target}} \otimes X_{s,j}^{\text{aux}} \otimes X_{s,(j+1) \bmod k/2}^{\text{aux}}.
$$

(D3)

Each perturbation term now acts on two target qubits instead of one, halving the order in perturbation theory at which the target Hamiltonian is recovered and would consequently only require $rk/2$ auxiliary qubits to obtain a result equivalent to Eq. (6).

In general, we can construct a gadget with an arbitrary target locality between three and k by acting on more target

qubits at once in each term of the perturbation. If the original target Hamiltonian is k body and the target locality of H^{gad} is k', the resulting gadget construction requires $k/(k'-2)$ auxiliary qubits for each term of the target Hamiltonian. Two of the k' operators of the tensor product have to be the pair of Pauli-X operators on the auxiliary register and the rest can be populated with the corresponding Pauli operators from h_s.

The statements above assume that $k/(k'-2)$ is equal to an integer; now let us look at the case where this is not the case. Let us define $k^{\top} = \lceil k/(k'-2) \rceil$ and $k^{\perp} = \lfloor k/(k'-2) \rfloor$. Using the insights from Appendix D 1, we can lift the divisibility requirement and can construct a k'-body gadget Hamiltonian with the perturbation given by

$$
\begin{aligned}
V_s = {} & \sum_{j=1}^{k^{\perp}} \tilde{c}_{s,j} \bigotimes_{\ell=1}^{k'-2} \sigma_{s,(j-1)(k'-2)+\ell}^{\text{target}} \otimes X_{s,j}^{\text{aux}} \otimes X_{s,(j+1)\bmod k^{\top}}^{\text{aux}} \\
& + \tilde{c}_{s,k^{\perp}+1} \bigotimes_{m=(k^{\perp}-1)(k'-2)+1}^{k} \sigma_{s,m}^{\text{target}} \otimes X_{s,k^{\perp}}^{\text{aux}} \otimes X_{s,(k^{\perp}+1)\bmod k^{\top}}^{\text{aux}}.
\end{aligned} \quad \text{(D4)}
$$

The last term may act on less than $k'-2$ target qubits but the paired X operator ensures that the only contribution is the one *acting* k times on the target register with k_s Pauli operators and $k - k_s$ identities.

When $k^{\perp} = k/(k'-2)$, the final term does not contribute and we recover the divisible case. Otherwise, the last term implements the remaining operators of the target term h_s. For a target Hamiltonian whose terms act nontrivially on k qubits simultaneously and a targeted locality of the gadget Hamiltonian of k', the number of required auxiliary qubits is $k^{\top} r$.

APPENDIX E: CAN PERTURBATIVE GADGETS HELP ADDRESS THE BARREN PLATEAU PROBLEM?

In this section, we provide an in-depth study of our results when applying perturbative gadgets in the context of gate-based quantum computing, in particular in the context of variational quantum algorithms. One might ask if and why one would even need a perturbative gadget without constraints of adiabaticity. After all, digital, gate-based quantum computing is not restricted by existing couplings due to the design of more intricate gate decompositions and the freedom of constructing arbitrary gates from universal gate sets. Although finding practical uses in general gate-based computations is as of yet an open problem, one fruitful application could be in the context of *variational quantum algorithms (VQAs)* [24–26], which are based upon measurements of a Hamiltonian expanded in the Pauli basis and for which it was shown that the Hamiltonian's locality, both in the sense of geometric locality and few-body terms, can play a role in their performance [47,48].

VQAs rely upon *parametrized quantum circuits* (PQCs) to evaluate a parameter-dependent cost function related to the expectation value of a set of observables and are greatly discussed in the setting of *noisy intermediate scale quantum (NISQ)* devices [27]. The probably simplest setting is that of the *variational quantum eigensolver* (VQE) [28], which refers to the case where the cost function is equal to the expectation

value of a Hamiltonian H^{target}, i.e.,

$$
C^{\text{target}}(\boldsymbol{\theta}) = \text{Tr}[H^{\text{target}} U(\boldsymbol{\theta}) \rho U(\boldsymbol{\theta})^{\dagger}], \quad \text{(E1)}
$$

for some initial state ρ, e.g., one given by the state vector $|00\cdots 0\rangle\langle 00\cdots 0|$, and quantum circuit ansatz $\boldsymbol{\theta} \mapsto U(\boldsymbol{\theta})$. Solving the problem corresponds to estimating the ground-state energy and associated ground-state vector of H^{target} by classically optimizing the parameters $\boldsymbol{\theta}$ and employing a quantum device to evaluate the aforementioned cost function.

A significant obstacle in the successful optimization is the presence of so-called *barren plateaus*, which reflect the phenomenon that the variance of gradients of the cost function with respect to the variational parameters decays exponentially in the number of qubits of the PQC. First discussed for the hardware-efficient ansatz with a polynomial number of layers in Ref. [30], the emergence of barren plateaus has been observed in other settings as well [12,31–33] and hinders the optimization procedure due to the therefore exponential number of required measurement samples. Scalable VQAs thus rely on practical mitigation strategies.

Existing approaches build upon improving initialization strategies [34], using correlated gates or restricted parameter spaces [32,35], intermediate measurements [36], the transferability of smooth solutions [37], layer-wise learning [38], pre-optimization [39,40], different types of ansatz [41], classical shadows [42], or alternative loss landscapes [43].

Furthermore, in Refs. [44–46], it has surprisingly been found that for large system sizes and circuit depths a *local* cost function, i.e., a cost function defined by a few-body Hamiltonian, results in more effective optimization compared with a *global* cost function, i.e., a cost function defined by a many-body Hamiltonian, and that global cost functions exhibit vanishing gradients for large system sizes already at constant circuit depths. The concept of such *cost-function-dependent* barren plateaus was formalized in Refs. [47,48], in which it has been further shown that local cost functions can still be optimized even for circuits of logarithmic depth.

Inspired by these findings, a natural question to ask is whether we can *localize* any given cost function, and thus avoid cost-function-dependent barren plateaus in the corresponding optimization landscape. For the problems considered in Refs. [44–46], the local cost function corresponding to the problem could be defined essentially by inspection of the global cost function. However, this is not necessarily possible in general and motivates us to open the discussion on whether perturbative gadgets could help extend results on the nonexistence of barren plateaus for local cost functions to nonlocal ones.

1. A reduction from k-body to three-body cost functions

Applying the above-presented gadget to substitute cost functions of the form in (E1), allows us to use Theorem 1 and Corollary 1 to conclude the following about VQAs: Given a cost function as in Eq. (E1), we can replace the k-body Hamiltonian H^{target} with the three-body gadget Hamiltonian H^{gad} from Definition 1, such that finding the minimum of the gadget Hamiltonian will bring us close to the low-energy subspace of the target Hamiltonian, and thus the minimum of the original cost function.
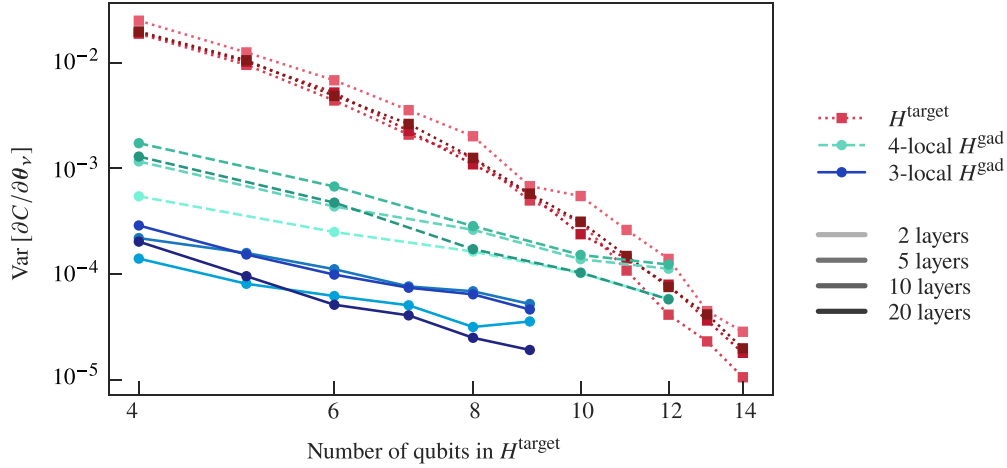
FIG. 7. Demonstration of the theoretical guarantee from Corollary 2. We take as our global cost function the one in Eq. (E1), defined by the target Hamiltonian $H^{\text{target}} = \bigotimes_{i=1}^{n} Z_i$, and we use an alternating layered ansatz with different circuit depths. Shown is the variance of the gradient of this global cost function (dark pink), the local cost function based on our three-body gadget Hamiltonian from Definition 1 (blue), and a similar local cost function defined by a four-body gadget Hamiltonian (turquoise) presented in Appendix D.

Furthermore, by *localizing* the original cost function, we obtain cost functions with provably nonvanishing gradients as shown in Refs. [47,48]. Specifically, they prove that the corresponding gradients decrease at most polynomially in the number of qubits for the same ansatz of a logarithmic depth. Thus, via Ref. [48, Theorem 2], we obtain the following result:

Corollary 2 (bound to the variance of the gradient of the cost function). Consider the local cost function $C^{\text{gad}}(\boldsymbol{\theta}) = \text{Tr}[H^{\text{gad}} U(\boldsymbol{\theta}) \rho U(\boldsymbol{\theta})^{\dagger}]$ defined by the gadget Hamiltonian in Definition 1, where $U(\boldsymbol{\theta})$ is the unitary corresponding to an alternating layered ansatz with L layers, with each block of the ansatz drawn independently from a local unitary 2-design. Then, the variance of the gradient of the cost function is bounded from below by

$$\text{Var}\left[\frac{\partial C^{\text{gad}}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_{\nu}}\right] \geqslant \Omega(\text{poly}(3^{-L})), \quad (E2)$$

where ν denotes the subset of parameters the gradient is evaluated for. We are thus guaranteed nonvanishing gradients for logarithmic circuit depths, i.e., for $L = O(\log n)$.

In Fig. 7, we present a proof-of-principle numerical demonstration of nonvanishing gradients. At this point, we would like to point out that simply having nonvanishing gradients does not imply successful optimization, i.e., trainability. Although the converse is true, one should, in general, be careful when implying trainability from the existence of gradients. Take, for example, the Hamiltonian $H = H^{\text{global}} \otimes \mathbb{1} + \mathbb{1}^{\otimes(n-1)} \otimes Z$. The vanishing gradients of the first, global term would effectively result in only the second term contributing to the gradients. That is, if H^{global} is a prototypical global observable afflicted by barren plateaus, it will barely contribute to the magnitude of the gradients. Consequently, the circuit will effectively be trained on the other, local, term $\mathbb{1}^{\otimes(n-1)} \otimes Z$, which has measurable gradients. In the end, one will successfully implement the gradient descent algorithm to a minimum, but only of the local part, and not of the full Hamiltonian. In other words, the total cost function has

measurable gradients, but these nonvanishing gradients do not help with the actual optimization problem.

Our gadget has gradients that do not vanish exponentially and a ground state that is close to the target solution. Unfortunately, it is not enough to fully avoid the core issue behind barren plateaus: the exponential increase in the number of required measurements. For the extreme case of $k = n$, the magnitude of the contribution of the target Hamiltonian in the effective Hamiltonian, or in other words the degeneracy splitting mentioned in Fig. 1, is suppressed exponentially by λ^n. For the results from perturbation theory to hold, λ has to be chosen to be small (see Theorem 1), further amplifying this problem. As a result, when reaching the low-energy subspace of the gadget Hamiltonian and therefore entering the regime described by the effective Hamiltonian, one would in practice need an exponentially increasing number of measurements to resolve the nuances in the cost function and find the global minimum. While we can guarantee a vanishing error in the limit of $\lambda \to 0$ and that all terms of the gadget Hamiltonian contribute to the gradient, the main contributions to the gradients would be essentially those corresponding to reaching the unperturbed ground space, i.e., some state vector $|\psi\rangle = |\phi\rangle \otimes |0\rangle^{\otimes rk}$. Therefore, actually finding the ground state of the gadget Hamiltonian might be prohibitively hard. However, this hardness can be present even for cost functions without barren plateaus, as discussed in detail in Refs. [52,53]. Moreover, for numerical proof-of-principle demonstrations, we find that leaving the strict regime of perturbation strengths corresponding to theoretical guarantees can help the optimization performance as shown in Appendix G.

For large λ we do not have such strong analytical guarantees but allow for larger contributions with respect to the perturbative terms. Our training simulations reflect these insights as we obtain different behaviors for different perturbation strengths. These observations are aligned with previous works on perturbative gadgets that have also found it beneficial to increase the perturbation strength, including increasing it to values outside the regime of convergence of perturbation theory [14,19]. Practical implementations might thus benefit

highly from treating the perturbation strength λ as a hyperparameter during optimization.

It is worth noting that while using the proposed gadget Hamiltonian requires rk additional qubits and comprises $2rk$ terms instead of r, there exists a simple, optimal measurement scheme for estimating its expectation value, relying on only four different measurement-basis settings, inspired by Refs. [54,55]. First, by measuring all auxiliary qubits in the Pauli-Z basis, we are able to estimate H^{aux}. Then, consecutively measuring all auxiliary qubits in the Pauli-X basis and simultaneously all target qubits in one of the three Pauli bases, we are able to estimate all terms V_s.

APPENDIX F: INAPPLICABILITY OF THE GADGET BY JORDAN AND FARHI FOR VARIATIONAL ALGORITHMS

The perturbative gadget proposed by Jordan and Farhi [5] relies strongly on the restriction of the allowed Hilbert space in which the state of the system can evolve. To be specific, in the language of Sec. IV, their gadget has a penalization Hamiltonian, as in Eq. (9), of the form $H = \sum_{1 \leqslant i_1 < i_2 \leqslant k} h_{i_1,i_2}$, with

$$h_{i_1,i_2} = \tfrac{1}{2}\left(\mathbb{1} - Z_{i_1} \otimes Z_{i_2}\right), \tag{F1}$$

and the perturbation operator A in Eq. (10) is such that

$$a_j = X_j, \quad j \in \{1, 2, \dots, k\}. \tag{F2}$$

However, since the ground space of their penalization Hamiltonian is two-dimensional and given by span$\{|0\rangle^{\otimes k}, |1\rangle^{\otimes k}\}$, they have to enforce that the state of the system remains in the $+1$ eigenspace of the $A = X^{\otimes k}$ operator, thereby reducing the system to a one-dimensional ground-state vector $|GS\rangle = |GHZ\rangle = (1/\sqrt{2})(|0\rangle^{\otimes k} + |1\rangle^{\otimes k})$. With this additional restriction, which can be fulfilled naturally in adiabatic quantum computing by initializing the auxiliary registers in Greenberger-Horne-Zeilinger (GHZ) states, it is easy to check that $X^{\otimes k}$ fulfills all properties stated in Eq. (11). Unfortunately, this restriction is not possible outside of the adiabatic regime, and their gadget construction is consequently not applicable to our problem at hand.

However, let us now consider what happens when one cannot use initialization and adiabatic restriction to ensure a reduction of the reachable Hilbert space. The unperturbed ground space of their penalization Hamiltonian H in Eq. (F1) is

$$\mathcal{E}^{(0)} = \mathrm{span}\left\{|\varphi\rangle^{\mathrm{target}} \bigotimes_{s=1}^{r} |\phi\rangle_s^{\mathrm{aux}} : |\varphi\rangle \in \mathcal{H}^{\mathrm{target}}, \right.$$
$$\left. |\phi\rangle \in \mathrm{span}\{|0\rangle^{\otimes k}, |1\rangle^{\otimes k}\}\right\}, \tag{F3}$$

and the corresponding projector is

$$P_0 = \mathbb{1} \otimes (|0\rangle\langle 0|^{\otimes k} + |1\rangle\langle 1|^{\otimes k})^{\otimes r} = \mathbb{1} \otimes \Pi_0^{\otimes r}. \tag{F4}$$

The fact that the dimension of $\mathcal{E}^{(0)}$ is not the same as the dimension of the target Hilbert space makes a considerable difference in Eq. (B31), where the shifted contributions become

$$\frac{\lambda^k}{\Xi} P_0 \left(\sum_{s=1}^{r} c_s h_s \otimes X_s^{\otimes k}\right) P_0. \tag{F5}$$

Using the fact that

$$\Pi_0 = |GHZ_+\rangle\langle GHZ_+| + |GHZ_-\rangle\langle GHZ_-|, \tag{F6}$$

with

$$|GHZ_\pm\rangle = \frac{1}{\sqrt{2}}(|0\rangle^{\otimes k} \pm |1\rangle^{\otimes k}), \tag{F7}$$

results in

$$\frac{\lambda^k}{\Xi} \sum_{s=1}^{r} c_s h_s \otimes \Pi_0^{\otimes s-1} \otimes (|GHZ_+\rangle\langle GHZ_+|$$
$$- |GHZ_-\rangle\langle GHZ_-|) \otimes \Pi_0^{\otimes r-s}. \tag{F8}$$

In other words, the effective Hamiltonian on the low-energy subspace of the gadget Hamiltonian is composed of a mixture of positive and negative contributions from each term as $\pm h_s$ is associated with different projectors on the different auxiliary registers. This prevents us from recovering the desired tensor-product structure between target and auxiliary registers with H^{target} required to use this construction in practice, for our purposes; see also Ref. [19, Definition 11].

APPENDIX G: NUMERICAL SIMULATIONS OF THE APPLICATION OF THE GADGET FOR VARIATIONAL QUANTUM ALGORITHMS

To show how one can use the proposed gadget construction, we have studied its performance on the toy model $H^{\mathrm{target}} = Z_1 Z_2 \cdots Z_n$. This specific example has already been used in the study of cost-function-dependent barren plateaus [32]. Furthermore, a single global Pauli string is the simplest example on which we can apply our gadget and that needs the least amount of resources. Indeed, we require $2n$ qubits to generate the corresponding cost function, while any additional global Hamiltonian term would imply n additional qubits. With this choice, having in mind the scaling of classical simulations, we can show the functioning of the gadget for the largest possible number of target qubits. The purpose of our simulation is to demonstrate Theorem 1 and Corollary 1; therefore, we performed exact, state-vector simulations, which are not subject to the considerations of exponential circuit evaluations presented previously. We used a layered ansatz as depicted in Fig. 8, as done in Refs. [30,32]. While this ansatz does not fulfill the exact conditions of Corollary 2, it allows a comparison with previous works on the topic.

1. Methods for gradient variance computations

As shown in Fig. 7, the variances of the gradients of the energy for the target Hamiltonian decay exponentially, while those of the gadget Hamiltonian seem to decay subexponentially. The exponential scaling is more obvious when using linear-log axes, but the current log-log choice puts emphasis on the polynomial scaling of the curves for the gadget cost functions. Indeed, with this choice of axes, a polynomial scaling results in a straight line which seems to fit the data points
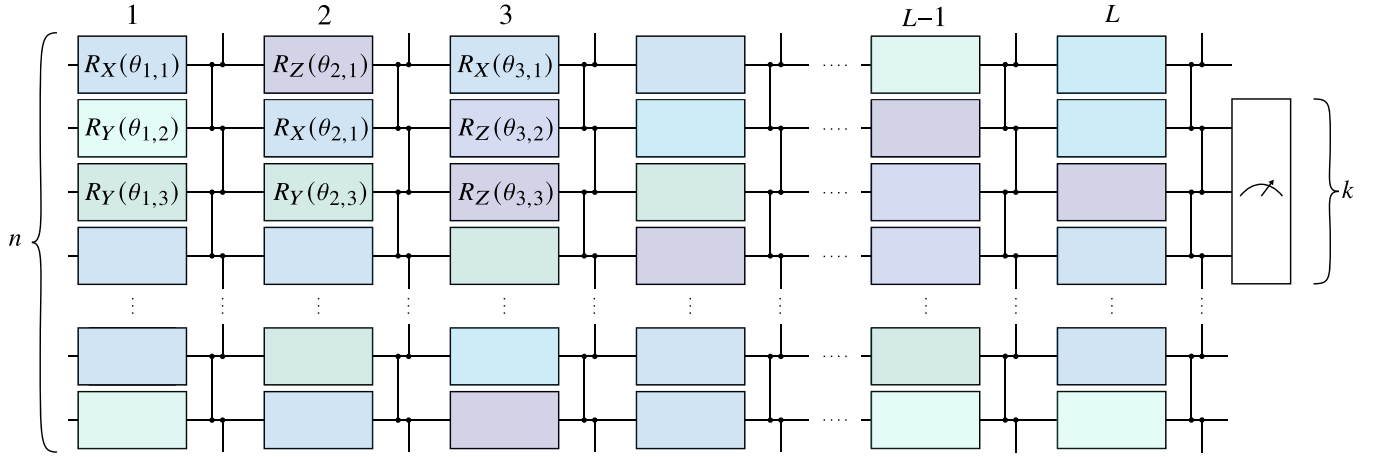
FIG. 8. Circuit ansatz used for the PQC implemented in the simulations of the use of our perturbative gadget for VQE. Each layer is composed of randomly chosen Pauli rotations on each qubit followed by controlled Pauli-Z gates in a chain pattern.

well. Although the presented regime does not show a clear advantage, the visible trend hints towards a crossover at larger qubit numbers that are out of reach for our classical simulator.

For these results, we have used the parameter on the nth qubit of the first layer when evaluating the variances. For the target Hamiltonian, this corresponds to the last qubit it acts upon, while it corresponds to the last qubit of the target register for the gadget Hamiltonian. Based on the asymmetry of the gadget Hamiltonian with respect to the target and auxiliary registers, choosing the parameter related to the last qubit of the target register within the first layer guarantees a contribution of all Hamiltonian terms at any depth. However, we stress that the found trend is expected to be similar for all parameter choices, as indicated by Corollary 2.

2. Training simulations

Additionally to the gradient variance simulations, we have simulated gradient-based optimization of H^{gad} while monitoring the expectation value of H^{target} for different values of

λ as shown in Fig. 9. From Theorem 1, we know that when successfully reaching the global minimum of the expectation value of the gadget Hamiltonian, we can expect the output state to converge to a state close to the ground state of H^{target}. It has to be pointed out, however, that the subtleties of the low-energy part of the spectrum are exponentially suppressed in λ, which can result in slow optimizations or even exponential requirements in the presence of noise. This can be seen in our experiments as the optimization stagnates for perturbation strengths within the range of validity of our theoretical results. Nevertheless, we have successfully obtained minimizations of the expectation value of the target Hamiltonian for large values of $\lambda > \lambda_{\max}$ within a reasonable number of iterations. We thus propose using the interaction strength λ as an additional hyperparameter and starting with larger values and reducing it later on for improved accuracy and faster training. Furthermore, we note that the qubit ordering has an impact on the practical performance and that a different ordering has been employed for the variances and training plots. Further details are discussed in Appendix G 3.
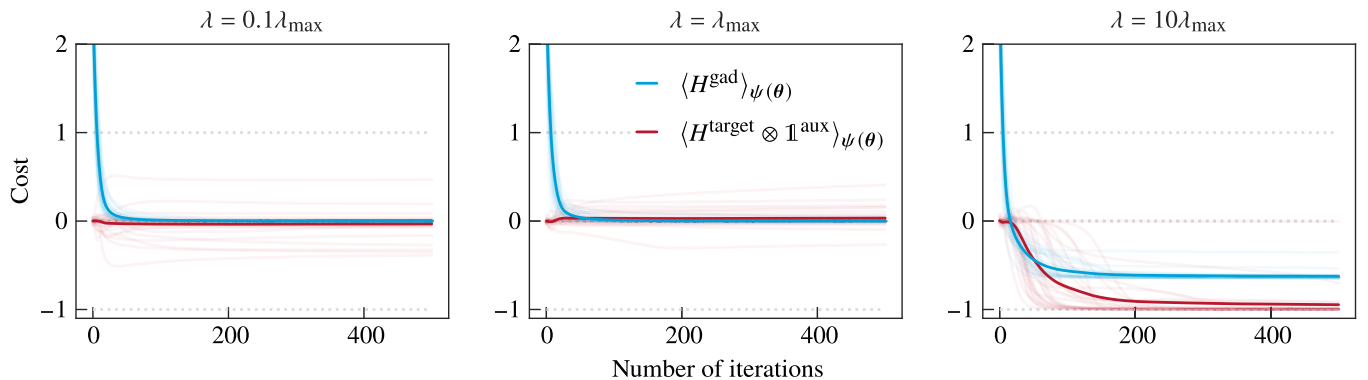

FIG. 9. Training simulations using the gadget construction H^{gad} introduced in Definition 1 as cost to optimize the global Hamiltonian $H^{\mathrm{target}} = Z^{\otimes 5}$. Each of the 30 runs is plotted in a light color while the more dark lines are the average over all runs. We employ an alternating layered ansatz on ten layers that is trained on H^{gad} (blue). Furthermore, we evaluate the energy of $H^{\mathrm{target}} \otimes \mathbb{1}$ (red) by measuring the expectation value of the target Hamiltonian for the output state of the circuit trained on the gadget Hamiltonian, hence tracing out the auxiliary part. While the theoretical guarantees of Theorem 1 only hold for $\lambda \leqslant \lambda_{\max}$, this bound is not tight and we observe improved training for larger values. We thus propose treating λ as an additional hyperparameter.
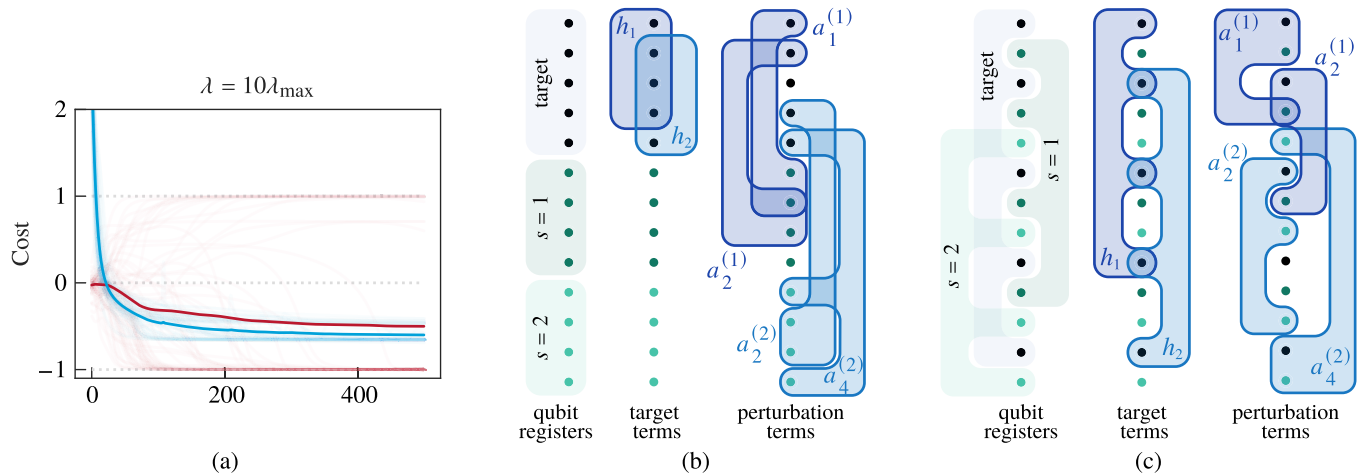
FIG. 10. Two possible qubit orderings and the resulting simulations for the case of $n = 5$, $k = 4$, and $r = 2$ (as in Fig. 5). Panel (a) is the corresponding simulation to Fig. 9 but using the basic ordering. Panels (b) and (c) show said ordering (used in Fig. 7) and the improved one (used in Fig. 9), respectively. For each, first is the distribution of the 13 qubits into the three registers (left), then the terms from the target Hamiltonian and the qubits they act on (center), and finally a selection of the terms composing the gadget Hamiltonian (right). In contrast with Fig. 5, we abandoned the circular layout and use a linear one to match the usual representation of the qubit register in a quantum computing circuit.

Note that these training simulations are to be seen as proof-of-principle demonstrations, since, in the simulated regime, even the original, global Hamiltonian can be optimized. Also, since the proposed gadget requires several times the number of qubits that the global Hamiltonian would, we quickly escape the realm of system sizes that can be efficiently simulated classically. Still, the scalings of our gadget construction remain compatible with the NISQ regime and could be a tool added to the arsenal of techniques for optimizing cost functions on near-term devices that are otherwise plagued by the barren plateau phenomenon.

3. Performance improvements through qubit reordering

Until this point, we focused more on the conceptual message of using perturbative gadgets in VQAs to overcome cost-function-dependent barren plateaus and less on practical details relevant to their actual implementation. One of them is the qubit ordering in the joint register of the quantum computer or simulator. While our main theorems do not depend on the geometry of the quantum circuit, the couplings in the constructed gadget Hamiltonian do have a particular structure. Considering the state generated at the end of the PQC and the nearest-neighbor coupling of each layer of the used ansatz in combination with shallow circuit depths, it is reasonable to assume that the ordering of the qubits in a one-dimensional, joint register will have an effect. Qubits coupled by the target Hamiltonian should probably be more entangled than those that are not coupled. We have thus explored different layouts of the target and auxiliary qubit registers with the goal of keeping coupled qubits closer.

First, let us discuss the straightforward qubit ordering used for the simulation results in Fig. 7. As shown in Fig. 10(b), the qubits have been added to the total register as they appear in the gadget construction. That is, first the target register and then one auxiliary register after the other. The resulting coupling terms $\sigma_{s,j}^{\mathrm{target}} \otimes X_{s,j}^{\mathrm{aux}} \otimes X_{s,(j+1) \bmod k}^{\mathrm{aux}}$ are then divided with

the target qubit being far from the two corresponding auxiliary qubits. Training simulations in this setting [Fig. 10(a)] were mostly successful, but in some instances ended up maximizing the target cost function instead of minimizing it.

Since the coupling graph resulting from H^{gad} does not allow for a one-dimensional distribution such that only nearest-neighbor qubits are coupled, we settled for putting at least one of the relevant auxiliary qubits close to the target qubit. To do so, we interleave the registers such that the first auxiliary qubit in $\sigma_{s,j}^{\mathrm{target}} \otimes X_{s,j}^{\mathrm{aux}} \otimes X_{s,(j+1) \bmod k}^{\mathrm{aux}}$ is added between the corresponding target qubit and the next target qubit. Due to the cyclic nature of the gadget construction, the second auxiliary qubit will be placed in the neighborhood of the following target qubit. A simple example of this is presented in Fig. 10(c). With this new ordering, we obtained the results shown in Fig. 9. The comparison of the convergence properties of the training simulations in both settings provides numerical evidence of the role of qubit ordering when it comes to practical implementations of the introduced methods. First, the average convergence speed visibly improves when using the new ordering. Second, the better ordering does not exhibit any runs that maximize the target Hamiltonian as found for the first one, which could be attributed to the local minima of the gadget cost function. Although this reordering does not pretend to be optimal but rather a first proposal, we could show for our example that considering the order of the qubits in combination with the structure of the minimized cost function can lead to clear improvements in the performance of VQAs in practical settings.

APPENDIX H: A GADGET FOR REDUCING MEASUREMENT BASES IN THE HAMILTONIAN READOUT PROBLEM

The gadget discussed so far is a gadget tuned for maximally reducing the locality of the gadget Hamiltonian. There, we have also noted, that the energy of such a gadget construction

can be estimated using only four measurement settings. This motivates the construction of a different gadget minimizing the resource overheads. Such a gadget is presented here. Furthermore, since such a gadget can be expressed in four layers of mutually commuting terms, it could also be applicable to the fault-tolerant setting, where layers of mutually commuting T gates can be executed in parallel by increasing the number of qubits [56].

1. Gadget definition

The gadget itself is constructed as follows: Let H^{target} be an arbitrary Hamiltonian represented using r differently weighted tensor products of Pauli operators, such that

$$H^{\text{target}} = \sum_{s=1}^{r} c_s H_s = \sum_{s=1}^{r} c_s H_s^X H_s^Y H_s^Z, \quad \text{(H1)}$$

where each individual Pauli-word is decomposed into its Pauli-X, Pauli-Y, and Pauli-Z part, e.g.,

$$H_1 = X_1 Y_2 X_3 Z_4, \quad \text{(H2)}$$

$$H_1^X = X_1 X_3, \quad \text{(H3)}$$

$$H_1^Y = Y_2, \quad \text{(H4)}$$

$$H_1^Z = Z_4. \quad \text{(H5)}$$

Then, we choose our gadget Hamiltonian to be

$$H^{\text{gad}} = H^{\text{aux}} + \lambda V, \quad \text{(H6)}$$

with

$$H^{\text{aux}} = -\sum_{i=1}^{q} Z_i \quad \text{(H7)}$$

acting as the unperturbed Hamiltonian on the q auxiliary qubits and

$$\lambda V = \lambda \sum_{s=1}^{r} \sum_{A \in \{X,Y,Z\}} \tilde{c}_s H_s^A \otimes V_{s,A}^{\text{aux}} \quad \text{(H8)}$$

its perturbation. Since the goal is that products of $V_{s,A}^{\text{aux}}$ only become the identity if the corresponding product of H_s^A is one of the target terms, we introduce

$$\tau_i^X = X_{i-1}, \quad \tau_i^Y = X_i, \quad \tau_i^Z = X_{i-1}X_i \quad \text{(H9)}$$

for X being the Pauli-X operators acting on the auxiliary qubits.

The idea is to construct the terms $V_{s,A}^{\text{aux}}$ as products of some τ_i^A such that their product only results in the identity if we have $V_s^X V_s^Y V_s^Z$. To do so, we define M to be the set of all subsets of $\{1, 2, \ldots, q/2\}$ with an odd number of entries sorted with respect to their cardinality. There are

$$\sum_{k \text{ odd}}^{q} \binom{q}{k} = 2^{q-1} \quad \text{(H10)}$$

such sets. Then, constructing the perturbations $V_{s,A}^{\text{aux}}$ for the term H_s^A using the set M_s, we define

$$V_{s,A}^{\text{aux}} = \prod_{i \in M_s} \tau_{2i}^A, \quad \text{(H11)}$$

and

$$\tilde{c}_s = \sqrt[3]{c_s |M_s|^2}. \quad \text{(H12)}$$

The ground state of H^{aux} is given by $|0\rangle^{\otimes k}$. Since the action of the perturbation is not immediately apparent, it is worthwhile to inspect it more closely. Imagine a system with $q = 10$ auxiliary qubits $a_1, \ldots, a_q$. There, the individual $V_{i,A}^{\text{aux}}$ are chosen in the fashion of Table II:

In this fashion, we have that $V_s^X V_{s'}^Y V_{s''}^Z = \mathbb{1}$ only if $s = s' = s''$. In any other case, due to the perturbative expansion, the product does not contribute to the effective Hamiltonian. Since there are two auxiliary qubits required as a unit, we obtain a qubit overhead of $q \geqslant 2\lceil \log_2(r) + 1 \rceil$. Here, we note that this perturbation structure is only one example and that more compact versions requiring fewer qubits might exist.

2. Mathematical analysis

The perturbation of H^{aux} by λV results in a global energy shift and a splitting of the degeneracy of the ground space, mimicking H^{target} and leading to the following result:

Theorem 2 (mathematical analysis). Given H^{gad} acting on n target and q auxiliary qubits as defined above and $\lambda < 1/2\|V\|$ with the operator norm $\| \ldots \|$, $H_{\text{eff}}(H^{\text{gad}}, d)$ will mimic H^{target} at third order in perturbation theory with respect to λ, specifically

$$H_{\text{eff}}(H^{\text{gad}}, 2^n) = \Delta \Pi + \lambda^3 H^{\text{target}} \otimes P_+ + O(\lambda^4), \quad \text{(H13)}$$

where Δ is a constant energy shift and $P_+ = |0\rangle\langle 0|^{\otimes q}$ denotes the operator projecting all auxiliary qubits onto their ground state and Π the projector on the support of $H_{\text{eff}}(H^{\text{gad}}, 2^k)$.

Proof. The fundamentals of the perturbative expansion are presented above in Appendix A, identical to those also used

TABLE II. Exemplary assignments for the perturbation acting on ten auxiliary qubits for a computational Hamiltonian comprising 16 terms.

	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}
$V_{1,X}^{\text{aux}} = \tau_2^X$	X	$\mathbb{1}$								
$V_{1,Y}^{\text{aux}} = \tau_2^Y$	$\mathbb{1}$	X								
$V_{1,Z}^{\text{aux}} = \tau_2^Z$	X	X								
$V_{2,X}^{\text{aux}} = \tau_4^X$			X	$\mathbb{1}$						
$V_{3,X}^{\text{aux}} = \tau_6^X$					X	$\mathbb{1}$				
$\vdots$						$\ddots$				
$V_{6,X}^{\text{aux}} = \tau_2^X \tau_4^X \tau_6^X$	X	$\mathbb{1}$	X	$\mathbb{1}$	X	$\mathbb{1}$				
$V_{7,X}^{\text{aux}} = \tau_2^X \tau_4^X \tau_8^X$	X	$\mathbb{1}$	X	$\mathbb{1}$			X	$\mathbb{1}$		
$\vdots$								$\ddots$		
$V_{16,X}^{\text{aux}} = \prod_{i=1}^{5} \tau_{2i}^X$	X	$\mathbb{1}$	X	$\mathbb{1}$	X	$\mathbb{1}$	X	$\mathbb{1}$	X	$\mathbb{1}$

in Appendix B. Consequently, the expansion converges if

$$\|\lambda V\| < \frac{\gamma}{4}, \tag{H14}$$

with γ being equal to the smallest energy gap of H^{aux}, i.e., $\gamma = 2$ for H^{gad} leading to the condition that

$$\lambda \leqslant \frac{1}{2\|V\|}. \tag{H15}$$

Let us now start by identifying nonzero terms in every order of perturbation theory. Since [see Eq. (A54)] all terms are sandwiched by $P_0 = \mathbb{1} \otimes P_+$ operators projecting the system onto the ground space $\mathcal{E}^{(0)}$ of H^{aux}, all terms vanish that do not take a state from $\mathcal{E}^{(0)}$ and return it to $\mathcal{E}^{(0)}$. For the perturbation introduced in Eq. (H8), this is equivalent to saying that no combination of terms of V resulting in an auxiliary space unequal to $|0\rangle^{\otimes q}$ will survive. Consequently, there will be no surviving terms in the first order of perturbation theory.

The terms in second order originate from V^2 and belong to one of the following two categories: either they vanish because they excite the auxiliary space in two different ways not returning the state to $\mathcal{E}^{(0)}$ or they are squares of individual terms that survive since Pauli operators square to the identity. The latter are responsible for the constant energy shift Δ.

In analogy, all third-order terms vanish except for those of the form $V_s^X V_s^Y V_s^Z$ resulting in a degeneracy splitting caused by the application of $H_s^X H_s^Y H_s^Z$ on the original system, because only a combination of τ_s^X, τ_s^Y, and τ_s^Z will yield $|0\rangle^{\otimes q} \in \mathcal{E}^{(0)}$ on the auxiliary register after the perturbation. There are six possible permutations of these three perturbations, and the perturbation corresponding to the Pauli-Z string can happen first, second or last, resulting in different energy penalties. Applying V_i^Z first or last will result in one intermediate state with a penalty of $4|M_s|$ and one with a penalty of $2|M_s|$, whereas applying the V_i^Z perturbation in the middle results in both intermediate states having a penalty of $2|M_s|$. Further-

more, from the sum in Eq. (A54), only the case $(\ell_1, \ell_2) = (1, 1)$ contributes, and not $(2, 0)$. Including the corresponding combinatorial factors we find

$$H_{\text{eff}}(H^{\text{gad}}, 2^n)^{(3)}$$

$$= \sum_{s=1}^{r} \frac{1}{|M_s|^2} \left(2 \underbrace{\frac{1}{(-2)(-2)}}_{V_i^Z \text{ second}} + 4 \underbrace{\frac{1}{(-4)(-2)}}_{V_i^Z \text{ first or last}} \right)$$

$$\times \tilde{c}_s^3 \lambda^3 H_s^X H_s^Y H_s^Z \otimes P_+$$

$$= \lambda^3 \sum_{s=1}^{r} c_s H_s \otimes P_+ = \lambda^3 H^{\text{target}} \otimes P_+. \tag{H16}$$

∎

The striking feature of these types of gadgets is, that their Hamiltonian can be grouped into four sums of qubit-wise commuting Pauli operators, which can therefore be estimated using only four measurement bases:

$$H^{\text{gad}} = -\underbrace{\sum_{i=1}^{q} \mathbb{1} \otimes Z_i}_{\mathbb{1}_{\text{target}} \otimes Z_{\text{aux}}} + \lambda \underbrace{\sum_{s=1}^{r} \tilde{c}_s H_s^X \otimes V_{s,X}^{\text{aux}}}_{X_{\text{target}} \otimes X_{\text{aux}}}$$

$$+ \lambda \underbrace{\sum_{s=1}^{r} \tilde{c}_s H_s^Y \otimes V_{s,Y}^{\text{aux}}}_{Y_{\text{target}} \otimes X_{\text{aux}}} + \lambda \underbrace{\sum_{s=1}^{r} \tilde{c}_s H_s^Z \otimes V_{s,Z}^{\text{aux}}}_{Z_{\text{target}} \otimes X_{\text{aux}}}. \tag{H17}$$

However, as mentioned above, the scaling in λ is unfavorable except for settings in which it is excessively expensive to change measurement settings but cheap to evaluate a circuit given a chosen setting.

---

[1] A. Kitaev, A. Shen, and M. Vyalyi, *Classical and Quantum Computation*, Graduate Studies in Mathematics Vol. 47 (American Mathematical Society, Providence, Rhode Island, 2002).

[2] J. Kempe and O. Regev, 3-local Hamiltonian is QMA-complete, Quantum Inf. Comput. **3**, 258 (2003).

[3] J. Kempe, A. Kitaev, and O. Regev, The complexity of the local Hamiltonian problem, SIAM J. Comput. **35**, 1070 (2006).

[4] R. Oliveira and B. M. Terhal, The complexity of quantum spin systems on a two-dimensional square lattice, Quantum Inf. Comput. **8**, 900 (2008).

[5] S. P. Jordan and E. Farhi, Perturbative gadgets at arbitrary orders, Phys. Rev. A **77**, 062329 (2008).

[6] T. S. Cubitt, A. Montanaro, and S. Piddock, Universal quantum Hamiltonians, Proc. Natl. Acad. Sci. USA **115**, 9497 (2018).

[7] S. Bravyi and M. Hastings, On complexity of the quantum Ising model, Commun. Math. Phys. **349**, 1 (2017).

[8] L. Zhou and D. Aharonov, Strongly universal Hamiltonian simulators, arXiv:2102.02991.

[9] S. Piddock and A. Montanaro, Universal qudit Hamiltonians, arXiv:1802.07130.

[10] S. Piddock and J. Bausch, Universal translationally-invariant Hamiltonians, arXiv:2001.08050.

[11] D. Harley, I. Datta, F. R. Klausen, A. Bluhm, D. S. França, A. H. Werner, and M. Christandl, Going beyond gadgets: The importance of scalability for analogue quantum simulators, arXiv:2306.13739.

[12] S. Wang, E. Fontana, M. Cerezo, K. Sharma, A. Sone, L. Cincio, and P. J. Coles, Noise-induced barren plateaus in variational quantum algorithms, Nat. Commun. **12**, 6961 (2021).

[13] J. D. Biamonte and P. J. Love, Realizable Hamiltonians for universal adiabatic quantum computers, Phys. Rev. A **78**, 012352 (2008).

[14] S. Bravyi, D. P. DiVincenzo, D. Loss, and B. M. Terhal, Quantum simulation of many-body Hamiltonians using perturbation theory with bounded-strength interactions, Phys. Rev. Lett. **101**, 070503 (2008).

[15] Y. Cao, R. Babbush, J. Biamonte, and S. Kais, Hamiltonian gadgets with reduced resource requirements, Phys. Rev. A **91**, 012315 (2015).

[16] Y. Cao, Combinatorial algorithms for perturbation theory and application on quantum computing, Ph.D. thesis, Purdue University, 2016.

[17] Y. Cao and S. Kais, Efficient optimization of perturbative gadgets, Quantum Inf. Comput. **17**, 779 (2017).

[18] Y. Subaşı and C. Jarzynski, Nonperturbative embedding for highly nonlocal Hamiltonians, Phys. Rev. A **94**, 012342 (2016).

[19] J. Bausch, Perturbation gadgets: Arbitrary energy scales from a single strong interaction, Ann. Henri Poincaré **21**, 81 (2020).

[20] C. G. Brell, S. T. Flammia, S. D. Bartlett, and A. C. Doherty, Toric codes and quantum doubles from two-body Hamiltonians, New J. Phys. **13**, 053039 (2011).

[21] S. A. Ocko and B. Yoshida, Nonperturbative gadget for topological quantum codes, Phys. Rev. Lett. **107**, 250502 (2011).

[22] C. Bloch, Sur la théorie des perturbations des états liés, Nucl. Phys. **6**, 329 (1958).

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

[24] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and P. J. Coles, Variational quantum algorithms, Nat. Rev. Phys. **3**, 625 (2021).

[25] K. Bharti, A. Cervera-Lierta, T. H. Kyaw, T. Haug, S. Alperin-Lea, A. Anand, M. Degroote, H. Heimonen, J. S. Kottmann, T. Menke, W.-K. Mok, S. Sim, L.-C. Kwek, and A. Aspuru-Guzik, Noisy intermediate-scale quantum algorithms, Rev. Mod. Phys. **94**, 015004 (2022).

[26] J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik, The theory of variational hybrid quantum-classical algorithms, New J. Phys. **18**, 023023 (2016).

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

[28] 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**, 4213 (2014).

[29] E. Farhi, J. Goldstone, and S. Gutmann, A quantum approximate optimization algorithm, arXiv:1411.4028.

[30] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Barren plateaus in quantum neural network training landscapes, Nat. Commun. **9**, 4812 (2018).

[31] C. Ortiz Marrero, M. Kieferová, and N. Wiebe, Entanglement-induced barren plateaus, PRX Quantum **2**, 040316 (2021).

[32] Z. Holmes, K. Sharma, M. Cerezo, and P. J. Coles, Connecting ansatz expressibility to gradient magnitudes and barren plateaus, PRX Quantum **3**, 010313 (2022).

[33] E. C. Martín, K. Plekhanov, and M. Lubasch, Barren plateaus in quantum tensor network optimization, Quantum **7**, 974 (2023).

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

[35] T. Volkoff and P. J. Coles, Large gradients via correlation in random parameterized quantum circuits, Quantum Sci. Technol. **6**, 025008 (2021).

[36] R. Wiersema, C. Zhou, J. F. Carrasquilla, and Y. B. Kim, Measurement-induced entanglement phase transitions in variational quantum circuits, SciPost Phys. **14**, 147 (2023).

[37] A. A. Mele, G. B. Mbeng, G. E. Santoro, M. Collura, and P. Torta, Avoiding barren plateaus via transferability of smooth

solutions in a Hamiltonian variational ansatz, Phys. Rev. A **106**, L060401 (2022).

[38] A. Skolik, J. R. McClean, M. Mohseni, P. van der Smagt, and M. Leib, Layerwise learning for quantum neural networks, Quantum Mach. Intell. **3**, 5 (2021).

[39] M. Kieferova, O. M. Carlos, and N. Wiebe, Quantum generative training using Rényi divergences, arXiv:2106.09567.

[40] J. Dborin, F. Barratt, V. Wimalaweera, L. Wright, and A. G. Green, Matrix product state pre-training for quantum machine learning, Quantum Sci. Technol. **7**, 035014 (2022).

[41] X. Liu, G. Liu, J. Huang, and X. Wang, Mitigating barren plateaus of variational quantum eigensolvers, arXiv:2205.13539.

[42] S. H. Sack, R. A. Medina, A. A. Michailidis, R. Kueng, and M. Serbyn, Avoiding barren plateaus using classical shadows, PRX Quantum **3**, 020365 (2022).

[43] B. T. Kiani, G. D. Palma, M. Marvian, Z.-W. Liu, and S. Lloyd, Learning quantum data with the quantum earth mover's distance, Quantum Sci. Technol. **7**, 045002 (2022).

[44] S. Khatri, R. LaRose, A. Poremba, L. Cincio, A. T. Sornborger, and P. J. Coles, Quantum-assisted quantum compiling, Quantum **3**, 140 (2019).

[45] R. LaRose, A. Tikku, E. O'Neel-Judy, L. Cincio, and P. J. Coles, Variational quantum state diagonalization, npj Quantum Inf. **5**, 57 (2019).

[46] C. Bravo-Prieto, R. LaRose, M. Cerezo, Y. Subasi, L. Cincio, and P. J. Coles, Variational quantum linear solver, Quantum **7**, 1188 (2023).

[47] 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).

[48] A. V. Uvarov and J. D. Biamonte, On barren plateaus and cost function locality in variational quantum algorithms, J. Phys. A: Math. Theor. **54**, 245301 (2021).

[49] C. Wille, R. Egger, J. Eisert, and A. Altland, Simulating topological tensor networks with Majorana qubits, Phys. Rev. B **99**, 115117 (2019).

[50] V. Bergholm, J. Izaac, M. Schuld, C. Gogolin, S. Ahmed, V. Ajith, M. S. Alam, G. Alonso-Linaje, B. AkashNarayanan, A. Asadi, J. M. Arrazola, U. Azad, S. Banning, C. Blank, T. R. Bromley, B. A. Cordier, J. Ceroni, A. Delgado, O. Di Matteo, A. Dusko *et al.*, PennyLane: Automatic differentiation of hybrid quantum-classical computations, arXiv:1811.04968.

[51] https://github.com/SimonCichy/barren-gadgets.

[52] E. R. Anschuetz and B. T. Kiani, Quantum variational algorithms are swamped with traps, Nat. Commun. **13**, 7760 (2022).

[53] L. Bittel and M. Kliesch, Training variational quantum algorithms is NP-hard, Phys. Rev. Lett. **127**, 120502 (2021).

[54] A. Elben, S. T. Flammia, H.-Y. Huang, R. Kueng, J. Preskill, B. Vermersch, and P. Zoller, The randomized measurement toolbox, Nat. Rev. Phys. **5**, 9 (2023).

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

[56] D. Litinski, A game of surface codes: large-scale quantum computing with lattice surgery, Quantum **3**, 128 (2019).