EFFICIENT OPTIMIZATION OF PERTURBATIVE GADGETS

YUDONG CAO

Department of Computer Science, Purdue University, West Lafayette, IN 47906, USA
Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138

SABRE KAIS

Department of Chemistry, Purdue University, West Lafayette, IN 47906, USA
Qatar Energy and Environment Research Institute, HBKU, Doha, Qatar
Santa Fe Institute, 1399 Hyde Park Rd, Santa Fe, NM 87501, USA

Perturbative gadgets are general techniques for reducing many-body spin interactions to two-body ones using perturbation theory. This allows for potential realization of effective many-body interactions using more physically viable two-body ones. In parallel with prior work (arXiv:1311.2555 [quant-ph]), here we consider minimizing the physical resource required for implementing the gadgets initially proposed by Kempe, Kitaev and Regev (arXiv:quant-ph/0406180) and later generalized by Jordan and Farhi (arXiv:0802.1874v4). The main innovation of our result is a set of methods that efficiently compute tight upper bounds to errors in the perturbation theory. We show that in cases where the terms in the target Hamiltonian commute, the bounds produced by our algorithm are sharp for arbitrary order perturbation theory. We provide numerics which show orders of magnitudes improvement over gadget constructions based on trivial upper bounds for the error term in the perturbation series. We also discuss further improvement of our result by adopting the Schrieffer-Wolff formalism of perturbation theory and supplement our observation with numerical results.

1 Introduction

Quantum many-body interactions arise in a variety of contexts in quantum information and quantum computation, such as topological quantum computing [1, 2, 3, 4], measurement-based model of quantum computing [5, 6, 7, 8, 9, 10], adiabatic simulation of quantum chemistry [11], universal adiabatic quantum computation[a][15, 16], as well as constructions of circuit-to-Hamiltonian mapping for QMA-completeness [17, 18]. Given the broad range of applications

---

[a]We note that there are also several proposals [12, 13, 14] of universal adiabatic quantum computation that uses only simple two-body interactions, thus circumventing the need for many-body interactions. On the other hand, some of these proposals such as [13] also involve perturbation theory for which the error estimation algorithms of this paper may be useful for optimizing the parameters of these constructions as well.

for many-body interactions, it is then of great interest to simulate the behaviours of these many-body systems using experimental quantum systems. However, the current technologies for realizing controllable quantum interactions are limited to two-body interactions, implying a need for reducing many-body interactions to two-body ones. Such reduction boils down to constructing a two-body Hamiltonian whose low-lying eigenspace captures the eigenvectors and eigenvalues of the many-body "target" Hamiltonian. The technique of *perturbative gadgets* [18, 19, 20, 21] fulfills precisely this task.

The basic idea of perturbative gadget is that given a many-body "target" Hamiltonian H_{targ}, we construct a two-body "gadget" Hamiltonian $\tilde{H}$ of the form $H + V$ such that the low energy effective Hamiltonian of $\tilde{H}$ is arbitrarily close to H_{targ}. The gadget Hamiltonian $\tilde{H}$ acts on not only the Hilbert space of the target Hamiltonian but also an ancilla space. In other words we are embedding the spectrum of a given many-body Hamiltonian onto the low energy sector of a two-body Hamiltonian that acts on a larger Hilbert space. The effectiveness of such embedding is established by using perturbation theory for computing the low-energy effective interaction of $\tilde{H}$ and show that terms involving H_{targ} appear at the first few orders and the total contribution from the remaining terms in the infinite series amounts to a small quantity.

As useful as the perturbative gadgets have been in the study of the complexity of various types of physical systems [18, 19, 22, 23], the need for convergence in the perturbation series requires high variability in the coupling strengths that appear in the gadget Hamiltonian [24], which impose challenges for experimentally implementating the gadget Hamiltonians. Constructions that avoid using perturbation theory for reducing from many-body to two-body interactions have indeed been proposed [25, 26, 27]. However, as far as the authors are aware of, none of the non-perturbative constructions can be applied on general many-body Hamiltonians in the same way as their perturbative counterparts, in the sense that the non-perturbative constructions always assume that the Hamiltonian of the entire system must take certain form, while perturbative gadgets can be applied to reduce any subset of terms in a target Hamiltonian to two-body without concerning the form of the other terms in the Hamiltonian.

Here we consider minimizing variability in coupling strengths in the gadget Hamiltonians. This is important because it directly translates to reducing the physical resource required for experimentally implementing perturbative gadgets. Prior efforts [24] have optimized gadget constructions in [19, 20] for reducing many-body interactions to two-body. Here we are interested in the gadget construction due to Kempe, Kitaev, Regev [18] and later generalized by Jordan and Farhi [21]. The perturbative analysis of this construction is significantly more involved than constructions analyzed previously in [24]. However, it is of interest to us due to numerical evidence in [11] using direct diagonalization of target and gadget Hamiltonians which suggests that this construction requires less variable range of couplings than the constructions presented in [19, 20].

The technique for optimizing the gadgets presented in this work generalizes our previous work [24] for the gadgets in [19, 20] and applies the general framework presented in [28]. The main innovation of our result is an efficient method for finding tight upper bounds for the error in the perturbation series (*i.e.* the sum of terms from a specific finite order to infinity). By "efficient" we mean that suppose the gadget Hamiltonian acts on n qubits, our algorithm

finds a tight upper bound and sometimes the *exact* expression for perturbation terms at any order r in time $O(n^r)$, even though each term in the perturbative expansion is of dimension $O(2^n)$. Of course, the efficiency of our method heavily exploits the structure of the gadget Hamiltonian [18, 21] and does not necessarily hold for general perturbation theory on spin systems. However, in [28, 29] we argue that the assumptions needed for establishing efficiency may apply for a broader class of physical Hamiltonians than perturbative gadgets.

2   Perturbation theory and perturbative gadgets

The basic setting that we consider for perturbative analysis is a Hamiltonian $\tilde{H} = H + V$ where H is diagonal in the computational basis with an energy gap Δ between the ground space and the first excited space, and V is a perturbation that contains some non-zero off-diagonal elements. The main formalism that we use for extracting the low-energy effective Hamiltonian of $\tilde{H}$ is the well-known Feynman-Dyson series [30] based on self energy. There are various other formulations of perturbation theory such as Schrieffer-Wolff transformation [31, 32], Bloch expansion [33] and Rayleigh-Schrödinger perturbation theory (see for example [34, Ch. 17]). However, in the context of present work we focus on self-energy expansion from Feynman-Dyson series. In Section 5 we will apply Schrieffer-Wolff transformation onto the gadget Hamiltonians and show a connection between Feynman-Dyson series and Schrieffer-Wolff transformation.

Define the subspace spanned by eigenstates of H with energy lower than $\Delta/2$ as the *low-energy subspace* $\mathcal{L}_-$ and its orthogonal complement as the *high-energy subspace* $\mathcal{L}_+$. The projectors onto these subspaces are defined as Π_- and Π_+ respectively. We then introduce the notation for projections of any operator O onto the subspaces: $O_+ \equiv \Pi_+ O \Pi_+$, $O_- \equiv \Pi_- O \Pi_-$. $O_{-+} \equiv \Pi_- O \Pi_+$ and $O_{+-} \equiv \Pi_+ O \Pi_-$. The setup of self-energy expansion $\Sigma_-(z)$ requires the definition of operator valued resolvents $G(z) = (zI - H)^{-1}$ and $\tilde{G}(z) = (zI - \tilde{H})^{-1}$ where z is a scalar and I is the identity matrix. Then the expression for self energy can be written as $\Sigma_-(z) = zI_- - [\tilde{G}_-(z)]^{-1}$. Using Taylor expansion we have [18]

$$\begin{aligned} \Sigma_-(z) &= H_- + V_- + V_{-+}G_+V_{+-} + V_{-+}G_+V_+G_+V_{+-} + \cdots \\ &= H_0 + V_- + T_2 + T_3 + \cdots . \end{aligned} \tag{1}$$

Note that the r-th order term is simply a matrix product

$$T_r = V_{-+}G_+(V_+G_+)^{r-2}G_+V_{+-}, \tag{2}$$

which gives rise to our later discussion (Section 4) that interprets it as a sum of walks on a graph. The self energy expansion is useful because it approximates the low-energy sector of the perturbed Hamiltonian $\tilde{H}$. This is captured precisely in [18, Theorem 3], which we restate below. In the present paper our analyses involve mainly two types of operator norms, namely the 2-norm $\|A\|_2 = \max_{|\psi\rangle} \|A|\psi\rangle\|_2$, which is equal to the "spectral radius" for Hermitian operators, and ∞-norm $\|A\|_\infty = \max_i \sum_j |\langle i|A|j\rangle|$, the "maximum row sum", where $\{|i\rangle\}$ is the set of computational basis states.

Theorem 1 ([18], Theorem 3 restated) *Given a Hamiltonian $\tilde{H} = H + V$ with H having a spectral gap Δ between the ground space and the first excited subspace, suppose $\|V\|_2 \leq \Delta/2$. If there exists a Hamiltonian H_{eff} whose energies are contained in the interval*

[a, b] and some real constant $\epsilon > 0$ such that $a < b < \Delta/2 - \epsilon$ and for $z \in [a - \epsilon, b + \epsilon]$, we have

$$\|\Sigma_-(z) - H_{eff}\|_2 \leq \epsilon,$$

then the j-th eigenvalue $\tilde{\lambda}_j$ of $\tilde{H}_-$ and the corresponding j-th eigenvalue of H_{eff} differ by at most ϵ, for any appropriate range of j values.

Theorem 1 states that closeness in the operator norm between the self energy $\Sigma_-(z)$ and H_{eff} implies closeness in eigenvalues. In fact it also implies closeness in eigenvectors (see [18, Lemma 11]). Hence the entire operator H_{eff} is captured by the low energy sector of $\tilde{H}$. The basic idea of perturbative gadgets is that for a given many-body Hamiltonian H_{targ}, one constructs a two-body Hamiltonian $\tilde{H} = H + V$ with H and V matching the setting described before Equation 1, and the self energy expansion according to Equation 1 contains H_{targ} in its leading orders which constitutes the effective Hamiltonian H_{eff}, while the remaining terms can be bounded from above by ϵ. The gadget construction [18, 21] considered in this work reduces an arbitrary many-body Hamiltonian

$$H_{\text{targ}} = \sum_{i=1}^{m} c_i H_{\text{targ},i} \tag{3}$$

where each $H_{\text{targ},i} = \sigma_{i,1}\sigma_{i,2}\cdots\sigma_{i,k}$ is a k-body term with $\sigma_{i,j} \in \{X, Y, Z, I\}$ being the j-th single-qubit operator in the i-th term in H_{targ}, to two-body. Here X, Y, and Z are Pauli operators and I is the identity operator. The gadget Hamiltonian $\tilde{H} = H + V$ works by first introducing a register of k ancilla qubits for each k-body term $H_{\text{targ},i}$. Hence there are km ancillas in total, each of which can be labelled as (i, j) with $i \in [m]$ (we use $[x]$ to denote the set $\{1, \cdots, x\}$) being the index of the register and $j \in [k]$ being the index of the ancilla within the register. For each register i we then impose a Hamiltonian $H^{(i)}$ which ferromagnetically couples every pair of qubits in the register. Precisely, $H = \sum_{i=1}^{m} H^{(i)}$ is defined with each $H^{(i)}$ having the form

$$H^{(i)} = \sum_{1 \leq s < t \leq k} \frac{\Delta}{2(k-1)}(I - Z_{i,s}Z_{i,t}). \tag{4}$$

Here in Equation 4 the operator $Z_{i,j}$ acts on the ancilla qubit (i, j). Accordingly, the perturbation $V = \sum_{i=1}^{m} V^{(i)}$ consists of the terms that couple each register of ancillas with the corresponding qubits that the $\sigma_{i,j}$ terms act on:

$$V^{(i)} = \sum_{j=1}^{k} \lambda_{i,j}\sigma_{i,j} \otimes X_{i,j}. \tag{5}$$

Note that the gadget Hamiltonian considered here is slightly different from the original constructions [18, 21] in that the spectral gap Δ is introduced in $H^{(i)}$ in Equation 4 and $\lambda_{i,j}$ are coupling coefficients that are assigned such that the effective low energy Hamiltonian of $\tilde{H}$ calculated using perturbation theory in Equation 1 gives rise to the target Hamiltonian in Equation 3.

The spectrum of each $H^{(i)}$ is easy to find: the subspace of states with j qubits in $|1\rangle$ state has energy $E_j = \frac{j(k-j)}{k-1}\Delta$. The ground state subspace of each register of ancillas is $\mathcal{L}_-^{(i)} = \text{span}\{|0\rangle^{\otimes k}, |1\rangle^{\otimes k}\}$. The gap between the ground state subspace and the first excited

subspace is Δ. The perturbation terms $V^{(i)}$ break the degeneracy of $\mathcal{L}_-^{(i)}$ and are set up such that the perturbed subspace approximates the spectrum of H_{eff} closely. If one applies the self energy expansion (Equation 1) to the gadget Hamiltonian defined according to Equations 4 and 5, it is apparent that at any order $r \leq k$, T_r is proportional to projection Π_- onto $\mathcal{L}_-$, since the only r-step transitions under V that non-trivially contribute to T_r are the ones that start from $|0\rangle^{\otimes k}$ (resp. $|1\rangle^{\otimes k}$) and return to $|0\rangle^{\otimes k}$ (resp. $|1\rangle^{\otimes k}$). At the k-th order, if k is odd, then T_k consists of only a linear combination of $H_{\text{targ},i}$ terms, since only k-step transitions that goes from $|0\rangle^{\otimes k}$ (resp. $|1\rangle^{\otimes k}$) to $|0\rangle^{\otimes k}$ (resp. $|1\rangle^{\otimes k}$). If k is even, then T_k consists of terms proportional to $H_{\text{targ},i}$ as well as a term proportional to Π_-. Substituting the H and V in Equations 4 and 5 into Equation 1 leads to a self energy of the form

$$\Sigma_-(z) = \underbrace{\gamma\Pi_- + \sum_{i=1}^{m} H_{\text{targ},i} \otimes \Pi_{X,i}}_{H_{\text{eff}}} + T_{k+1} + T_{k+2} + \cdots \qquad (6)$$

where $\gamma = \gamma_1 + \cdots + \gamma_k$ is a scalar which sums over all contributions up to the k-th order and $\Pi_{X,i} = |0\rangle\langle 1|^{\otimes k} + |1\rangle\langle 0|^{\otimes k}$ acts on the i-th register, where we will show that each γ_i is efficiently computable by Algorithm 2 in Section 4.2. The remaining terms T_{k+1} and so on are error terms that should be suppressed to below ϵ by assigning Δ to be appropriately large.

3 Improving and optimizing gadget constructions

From the construction presented in the previous section, note that here the energy gap Δ is a crucial parameter that decides how accurate the perturbation theory is when applied to the gadget Hamiltonian $H + V$. The larger Δ is, the more accurately the gadget Hamiltonian captures the spectrum of the target Hamiltonian in its low energy subspace. However, larger values of Δ means more challenges for realizing the gadget Hamiltonian on an experimental system. This is because realizing the gadget Hamiltonian on a physical quantum system requires setting the coupling strengths of both the unperturbed Hamiltonian H, which is of the magnitude of Δ, and that of the perturbation V, whose strength could differ substantially from that of H. In other words, the requirement for variability in coupling strength becomes more stringent as Δ increases. For a system of n qubits with poly(n) many-body terms, typically Δ scales as poly(n) [19], which is unphysical for physical systems whose interactions are local.

The scaling of Δ as poly(n) is one of the main reasons for hesitation among researchers in using perturbative gadgets for reducing many-body interactions to two-body ones, thus motivating various non-perturbative constructions for special cases [25, 26, 27]. This situation can be remedied for perturbative gadgets by either insisting on assigning Δ independent of system size at a cost of *extensive* error $O(n\epsilon)$ [20], or substituting the current gadget construction with one that requires *arbitrarily weak* interaction strengths at a cost of poly(ϵ^{-1}) qubits [35]. Whichever gadget construction one wishes to adopt, there is always a practical optimization question: what is the *minimum* value of Δ such that the error does not exceed ϵ? In other words, what is the value of Δ for which the error $\|\Sigma_-(z) - H_{\text{eff}}\|_2$ is precisely ϵ? This is the question that we address in this paper, as minimizing Δ is essentially minimizing the physical resource needed for realizing the gadget Hamiltonian.
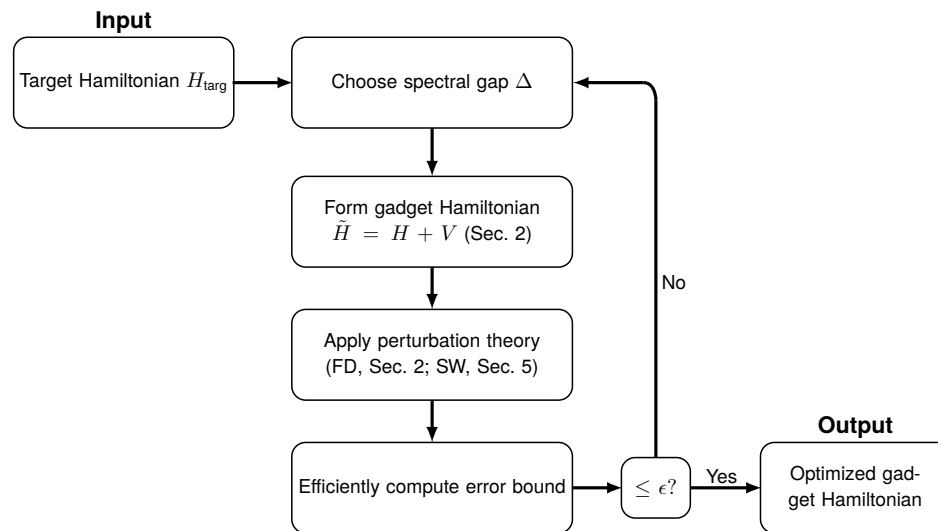
Fig. 1. The flow chart of optimizing the gadget Hamiltonian considered in this paper. We start by choosing a value Δ based on simple but loose error bounds that are dependent on ϵ, such as Equation 7. Then we construct the gadget Hamiltonian and use perturbation theory to find the effective Hamiltonian up to a certain order (FD stands for Feynman-Dyson series, which is introduced in Section 2 and SW stands for Schrieffer-Wolff transformation, which will be introduced in Section 5). The norm of terms from the order of Hamiltonian on to infinite order are bounded from above by efficiently computed error bounds. Techniques for producing tight error bounds without extensive computation is the central theme of the present paper.

In Figure 1 we present a flow chart of this optimization process. A challenge one has to face is then how to compute the error $\|\Sigma_-(z) - H_{\text{eff}}\|_2$, which is generally hard (on a classical computer) simply due to the exponential size of the Hilbert space as system size grows. We pursue a different strategy, which is to seek an upper bound of the form $\|\Sigma_-(z) - H_{\text{eff}}\|_2 \leq \|T_{k+1}\|_2 + \|T_{k+2}\|_2 + \cdots$, without requiring exponential-size computation in the number of qubits. In [24] we use the upper bound

$$\|T_r\|_2 = \|V_{-+}(G_+V_+)^{r-2}G_+V_{+-}\|_2 \leq \frac{1}{|z-\Delta|^{r-1}}\|V_{-+}\|_2^2 \cdot \|V_+\|_2^{r-2} \tag{7}$$

which appears to be tight [24, Figures 2 and 4] for the case where $m = 1$ *i.e.* there is only one target term to be reduced and one is using gadget constructions from [19]. However, when $m > 1$ and multiple gadgets are applied, the upper bound in Equation 7 becomes loose [24, Figure 3b]. This is because when perturbation theory of multiple ancilla qubits are concerned, the crude upper bound in Equation 7 could no longer capture the fine-grained details of the matrix product involved in T_r. Therefore we are also unable to use Equation 7 for finding tight error bounds for the gadget constructions presented in Section 2, since we are dealing with registers of ancillas of size $k > 1$.

We will show that by considering the more fine-grained details of matrix multiplication in T_r and exploiting the structure of the construction in Equations 4 and 5, it is possible to find a tight upper bound to $\|T_r\|_2$ in $O(m^r)$ time, which is polynomial in m for fixed r, thus enabling efficient optimization of Δ in the gadget construction.

As a final remark of the section, we note that it suffices to consider T_r for fixed r even though in the perturbation series r goes to infinity. Let $\lambda = \max_{i,j} \lambda_{i,j}$ with $\lambda_{i,j}$ defined in Equation 5. Our goal for the gadget construction is that at k-th order we have the target Hamiltonian in Equation 3 with coupling coefficients $c_i = O(1)$, while the terms at $(k+1)$-st order should be $O(\epsilon)$. This implies that roughly $\lambda^k/\Delta^{k-1} = O(1)$ and $\lambda^{k+1}/\Delta^k = O(\epsilon)$, which implies that $\Delta = O(\epsilon^{-k})$ and $\lambda = O(\epsilon^{-(k-1)})$. Hence $\|T_r\|_2 = O(\lambda^r/\Delta^{r-1}) = O(\epsilon^{r-k})$. Assuming that the locality of target Hamiltonian k is fixed, as r increases the norm of $\|T_r\|_2$ quickly becomes small enough to justify using the crude bound in Equation 7 on the remaining terms of the self-energy expansion. Therefore it suffices to consider r up to $k + d$ for some fixed d such that the total magnitude of the remaining sum is $O(\epsilon^d)$. For $r > k+d$, the upper bound becomes sufficiently small (assuming $\|T_r\|_\infty \to 0$ as $r \to \infty$), we use Equation 7 to bound the terms from $r = k + d + 1$ to infinity.

4 Efficiently computed tight error bound

In this section we present the details of our techniques. Section 4.1 introduces the notions that we use for reducing the amount of computation needed for the upper bound. These notions (such as "configuration" and "reduced configuration" as will be discussed in Section 4.1) are essentially simplified representations of H eigenstates. In Appendix 1 we provide an explicit example for calculating the norm of T_2 that illustrates the uses of these notions, hoping that the presentation be as instructive to the reader as possible. In Section 4.2 we present general algorithms for computing an upper bound to $\|T_r\|_2$ for any fixed r.

In Section 4.3 we prove that in the case where the terms $H_{\text{targ},i}$ pairwise commute, our algorithms in fact computes the *exact* value of $\|T_r\|_2$. We accomplish this by first developing

further properties of the notions that are introduced in Section 4.1, and then show how the algorithms introduced in Section 4.2 use these properties to effectively collect all the terms that contribute to $\|T_r\|$.

4.1 *Reducing the space of summation to polynomial size*

For simplicity from here on we let $\lambda_{i,j} = c_i^{1/k} \equiv \lambda_i$ in Equation 5 and define the vector $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \cdots, \lambda_m)$. Let $E^{(i)}(j) = \frac{j(k-j)}{k-1}\Delta \equiv E_j$ be the energy level of $H^{(i)}$ in Equation 4 with j out of k ancillas in $|1\rangle$ state. For a k-local target Hamiltonian of m terms the computational basis states of the km ancilla qubits can be represented by km-bit strings of the form $s_1 s_2 \cdots s_m$ with each s_i being a k-bit string describing the state of the ancilla qubits in the i-th register. For any $\phi \in \{0,1\}^{km}$, we use the notation $E(\phi) = \langle\phi|H|\phi\rangle$ to represent the energy of the ancilla state $|\phi\rangle$. Recall that $H = \sum_{i=1}^{m} H^{(i)}$ and the $H^{(i)}$ terms pairwise commute. As an explicit connection between the $E(\phi)$ and $E^{(i)}(j)$ notations, defining $h(s)$ as the Hamming weight of a string s, we have

$$E(\phi) = \sum_{i=1}^{m} E^{(i)}(h(s_i)). \tag{8}$$

In order to gain more insights about the structure of the r-th order term T_r in Equation 2, we insert resolutions of identity $I = \sum_{\phi \in \{0,1\}^{km}} |\phi\rangle\langle\phi|$ between the V and G operators in T_r. From Equation 2 we get

$$\begin{aligned} T_r &= \sum_{\phi_0, \phi_1, \cdots, \phi_r \in \{0,1\}^{km}} \left(\langle\phi_0|V|\phi_1\rangle \frac{1}{z - E(\phi_1)} \langle\phi_1|V|\phi_2\rangle \cdots \right. \\ &\qquad\qquad \left. \cdots \frac{1}{z - E(\phi_{r-1})} \langle\phi_{r-1}|V|\phi_r\rangle \right) \otimes |\phi_0\rangle\langle\phi_r|. \end{aligned} \tag{9}$$

Note that in the basis of $|\phi\rangle$ states, each matrix block $\langle\phi_0|T_r|\phi_r\rangle$ is a sum of 2^{kmr} terms, which is an enormous amount of computation. However, we note that a majority of summants are in fact zero. In order for a sequence of states $(\phi_0, \phi_1, \cdots, \phi_r)$ to have non-zero contribution to the sum in Equation 9, there are certain conditions that must be satisfied:

1. Because the self-energy $\Sigma_-(z)$ is restricted to the low energy subspace $\mathcal{L}_-$ of the unperturbed Hamiltonian H, both $|\phi_0\rangle$ and $|\phi_r\rangle$ must be restricted to $\mathcal{L}_- = \bigotimes_{i=1}^{m} \mathcal{L}_{-.i} = \text{span}\{|0^k\rangle, |1^k\rangle\}^{\otimes m}$. In other words, they must be of the form $s_1 s_2 \cdots s_m$ with each $s_i \in \{0^k, 1^k\}$;

2. Because each $V^{(i)}$ contains only single Pauli X operators acting on the ancilla qubits, for any pair of ancilla states $|\phi_i\rangle, |\phi_j\rangle$, the matrix block $\langle\phi_i|V|\phi_j\rangle$ is nonzero iff ϕ_i and ϕ_j differ by one and only one bit;

3. Because the V_+ and G_+ terms in T_r are projections onto the high energy subspace $\mathcal{L}_+$ of H, all the intermediate states ϕ_1 through ϕ_{r-1} must also belong to $\mathcal{L}_+$. In other words, they must *not* be of the form $s_1 s_2 \cdots s_m$ with each $s_i \in \{0^k, 1^k\}$.

Conditions 1 and 3 are reminiscent of Goldstone's theorem in quantum many-body physics [30, 36], where all non-zero contributions to the spectral difference between the perturbed

and the unperturbed systems is a summation of "connected diagrams" *i.e.* the state with no particles or holes present can never occur as an intermediate state because the resulting matrix element will contain disconnected parts. For condition 2 above if ϕ_i and ϕ_j differ at bit which belongs to the p-th register of k ancilla qubits, then

$$\|\langle\phi_i|V|\phi_j\rangle\|_\infty = \lambda_p. \tag{10}$$

Let $\mathcal{W}_r$ be the set of sequences $(\phi_0, \phi_1, \cdots, \phi_r)$ that satisfy the above three conditions. Then the summation in Equation 9 can be replaced by a summation over $\mathcal{W}_r$. The following Lemma states that to find an upper bound to $\|T_r\|_2$ from Equation 9 it suffices to consider a subset of sequences with *fixed* $\phi_0 = 0^{km}$, which is a string of km zeros. For convenience from here on we use "$\phi \in \mathcal{L}_-$" to as a shorthand for "$\phi \in \{0,1\}^{mk} : |\phi\rangle \in \mathcal{L}_-$".

Lemma 1 *For the r-th order term T_r as written in Equation 2, we have*

$$\|T_r\|_2 \leq \sum_{\phi_r \in \mathcal{L}_-} \|\langle 0^{mk}|T_r|\phi_r\rangle\|_\infty. \tag{11}$$

Proof. Let the operator $S_i = \bigotimes_{j=1}^k X_{i,j}$, recalling the subscript notation "i, j" means the j-th ancilla in the i-th register (Equation 5). Then for a state $|\phi\rangle$ of the mk ancillary qubits, S_i acting on $|\phi\rangle$ flips all the bits in the i-th register. From Equation 5 it is clear that for any sequence of S_i operations

$$S = S_{i_1} S_{i_2} \cdots S_{i_\ell} \tag{12}$$

with $i_1, i_2, \cdots, i_\ell \in [m]$, we have $SVS = V$. Additionally, the set of all ancilla states $|\phi\rangle \in \mathcal{L}_-$ with $\phi \in \{0,1\}^{mk}$ is invariant with respect to S_i for any i. Finally, substituting V with SVS in Equation 9 leads to

$$\langle\phi_0|ST_rS|\phi_r\rangle = \langle\phi_0|T_r|\phi_r\rangle. \tag{13}$$

Because we could express any $\phi \in \mathcal{L}_-$ as $|\phi\rangle = S|0^{mk}\rangle$ for some S with the form in Equation 12, we could further write T_r as

$$
\begin{aligned}
T_r &= \sum_{\phi_r \in \mathcal{L}_-} \sum_S \langle 0^{km}|ST_r|\phi_r\rangle S|0^{km}\rangle\langle\phi_r| \\
&= \sum_{\phi_r \in \mathcal{L}_-} \sum_S \langle 0^{km}|T_rS|\phi_r\rangle S|0^{km}\rangle\langle\phi_r| \\
&= \sum_{\phi_r \in \mathcal{L}_-} \langle 0^{km}|T_r|\phi_r\rangle \otimes \underbrace{\left(\sum_S S|0^{km}\rangle\langle\phi_r|S\right)}_{(*)}.
\end{aligned}
\tag{14}
$$

Here the summation $\sum_S$ is over all operators of the form in Equation 12. Going from the first line to the second we have used Equation 13. Going from the second line to the third is a substitution of variable $|\phi_r\rangle \rightarrow S|\phi_r\rangle$. Note that the term $(*)$ in Equation 14 is a sum of projectors that are orthogonal to each other (*i.e.* each pair multiply to zero), implying that the norm of $(*)$ is always one. Therefore to find an upper bound to $\|T_r\|_2$ from Equation 9 it suffices to consider a *fixed* $\phi_0 = 0^{km}$, which is a string of km zeros (For convenience from here on we use "$\phi \in \mathcal{L}_-$" to as a shorthand for "$\phi \in \{0,1\}^{mk} : |\phi\rangle \in \mathcal{L}_-$"):

$$\|T_r\|_2 \leq \|T_r\|_\infty \leq \max_{\phi_0 \in \mathcal{L}_-} \sum_{\phi_r \in \mathcal{L}_-} \|\langle\phi_0|T_r|\phi_r\rangle\|_\infty = \sum_{\phi_r \in \mathcal{L}_-} \|\langle 0^{mk}|T_r|\phi_r\rangle\|_\infty. \tag{15}$$

Here in Equation 15 the first $\leq$ uses $\|T_r\|_\infty = \|T_r\|_1$ implied by the hermiticity of T_r and the property $\|T_r\|_2^2 \leq \|T_r\|_1 \cdot \|T_r\|_\infty$, where $\|\cdot\|_1$ is the 1-norm of a matrix defined as "maximum column sum" $\|A\|_1 = \max_j \sum_i |\langle i|A|j\rangle|$. The second $\leq$ uses the definition of the ∞-norm. The final equality in Equation 15 comes from the invariance property described in Equation 14. $\qquad\square$

We further partition $\mathcal{W}_r$ into subsets according to different combinations of ϕ_0 and ϕ_r. Denote $W_r(\phi_0, \phi_r)$ as the subset of sequences in $\mathcal{W}_r$ that starts from ϕ_0 and ends at ϕ_r. For a given sequence $(\phi_0, \phi_1, \cdots, \phi_r)$, let $p_i \in [m]$ be the index of the ancilla register that contains the bit where ϕ_i differs from ϕ_{i-1}. Using Equation 10, the norm of each matrix block $\langle \phi_0|T_r|\phi_r\rangle$ can be bounded from above by

$$\|\langle \phi_0|T_r|\phi_r\rangle\|_\infty \leq \sum_{\mathcal{W}_r(\phi_0,\phi_r)} \underbrace{\lambda_{p_1} \cdot \frac{1}{|z - E(\phi_1)|} \cdot \lambda_{p_2} \cdots \lambda_{p_{r-1}} \cdot \frac{1}{|z - E(\phi_{r-1})|} \cdot \lambda_{p_r}}_{\equiv t_\phi(\phi_0, \cdots, \phi_r)} \tag{16}$$

where the weight function t_ϕ describes the contribution, or the "weight" of a specific sequence in the sum. Regardless of the restriction to $\mathcal{W}_r$, evaluating the upper bound in Equation 16 with brute-force enumeration of all possible intermediate steps $\phi_1, \cdots, \phi_{r-1}$ would still lead to a computational cost that is exponential in the number of registers m. However, by exploiting the structure of the gadget Hamiltonian we could reduce it to poly(m) for any fixed order r of perturbation theory. Such reduction is accomplished by introducing a sequence of two mappings $\mathbf{c}$ and $\tilde{\mathbf{c}}$ (Figure 2) which we will introduce in the following discussion.

Definition 1 (Configuration) *For a state $|\phi\rangle$ with $\phi = s_1 s_2 \cdots s_m$ where each s_i describes the state of a k-qubit register, we define the vector $\mathbf{c}(\phi) = (j_1, j_2, \cdots, j_m)$ with $j_i = h(s_i)$ as the* configuration *of a state $|\phi\rangle$. Each element j_i of the configuration corresponds to energy level $E^{(i)}(j_i)$ of the term $H^{(i)}$.*

Previously we have defined the set $\mathcal{W}_r$ as the set of r-step sequences $(\phi_0, \phi_1, \cdots, \phi_r)$ that contribute non-trivially to T_r. Since each bit string ϕ is associated with a configuration $\mathbf{c}$, each sequence of ancilla states (bit strings) is naturally associated with a sequence of configurations $(\mathbf{c}_0, \mathbf{c}_1, \cdots, \mathbf{c}_r)$. Similar to $\mathcal{W}_r$, we define $\mathcal{W}_r^{\mathbf{c}}$ as the set of r-step sequences $(\mathbf{c}_0, \mathbf{c}_1, \cdots, \mathbf{c}_r)$ that correspond to the r-step sequences in $\mathcal{W}_r$. We then let $\mathcal{W}_r^{\mathbf{c}}(\mathbf{c}_0, \mathbf{c}_r)$ be the set of r-step configuration sequences that starts from specific values of $\mathbf{c}_0$ and $\mathbf{c}_r$. With the Definition 1 we could rewrite the sum in Equation 16 as

$$\|\langle \phi_0|T_r|\phi_r\rangle\|_\infty \leq \sum_{\mathcal{W}_r^{\mathbf{c}}(\mathbf{c}_0,\mathbf{c}_r)} \sum_{\mathbf{c}(\phi_i)=\mathbf{c}_i} t_\phi(\phi_0, \cdots, \phi_r) = \sum_{\mathcal{W}_r^{\mathbf{c}}(\mathbf{c}_0,\mathbf{c}_r)} t_{\mathbf{c}}(\mathbf{c}_0, \cdots, \mathbf{c}_r) \tag{17}$$

where $t_{\mathbf{c}}(\mathbf{c}_0, \cdots, \mathbf{c}_r)$ is the total weight of all sequences $(\phi_0, \cdots, \phi_r)$ for which $\mathbf{c}(\phi_i) = \mathbf{c}_i$.

Note that unlike Equation 16 which sums over states that dwell in a space of dimension 2^{mk}, the summation in Equation 17 sums over configurations, which dwells in a space of dimension $O(k^m)$. This is an exponential reduction with respect to k, but the dimension is still exponential in m nonetheless. To reduce the dimension further, we first note that the energy of a configuration is invariant with respect to permutations of the ancilla registers. We would like to exploit this permutation invariance by restricting to a class of configuration vectors that are *sorted*. Specifically, we put forward the following definition.
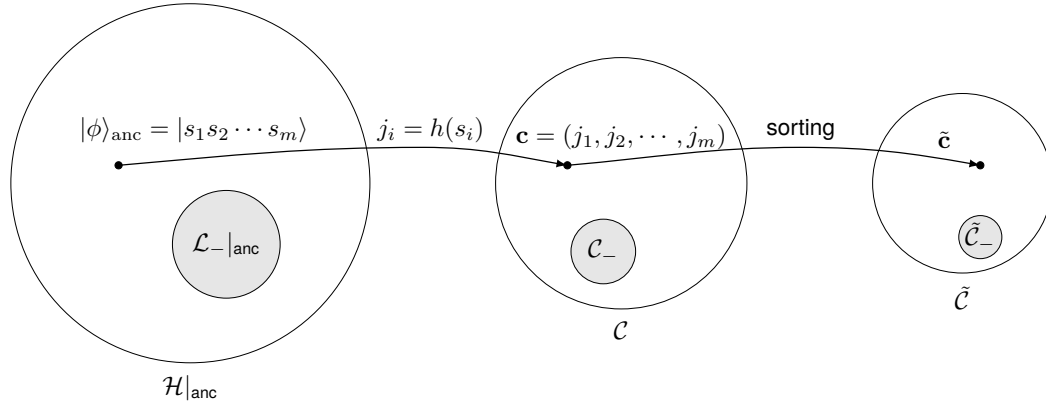
Fig. 2. Relationship between elements of the various spaces that are relevant to our discussion in Section 4.1. We use the notation $|_{\mathrm{anc}}$ to represent the restriction of the Hibert spaces on which the gadget Hamiltonian acts to the ancilla registers. Here the s_i's are k-bit strings, and $h(\cdot)$ is the Hamming weight of a string. We highlight the subspaces (or subsets) that correspond to the low energy subspace of the unperturbed Hamiltonian H. $\mathcal{C}_-$ consists of all the configurations of ancilla states in $\mathcal{L}_-|_{\mathrm{anc}}$ and $\tilde{\mathcal{C}}_-$ consists of the reduced configurations of those in $\mathcal{C}_-$.

Definition 2 (Reduced configuration) *For a configuration* $\mathbf{c}$ *as in Definition 1, we define the* reduced configuration *of* $\mathbf{c}$*, denoted as* $\tilde{\mathbf{c}}(\mathbf{c})$*, as the vector obtained by sorting a configuration* $\mathbf{c}$ *in non-decreasing order.*

Then by definition any set of configurations that differ by only permutations of the ancilla registers share the same reduced configurations. With an elementary inductive argument we can show that while there are $O(k^m)$ configurations, there are only $O(m^k)$ reduced configurations [37].

Since each sequence of configurations $(\mathbf{c}_0, \mathbf{c}_1, \cdots, \mathbf{c}_r)$ has a corresponding sequence of reduced configurations $(\tilde{\mathbf{c}}_0, \tilde{\mathbf{c}}_1, \cdots, \tilde{\mathbf{c}}_r)$, we define $\widetilde{\mathcal{W}_r^{\mathbf{c}}}$ as the set of reduced configuration sequences derived from all sequences of configurations in $\mathcal{W}_r^{\mathbf{c}}$. We then let $\widetilde{\mathcal{W}_r^{\mathbf{c}}}(\tilde{\mathbf{c}}_0, \tilde{\mathbf{c}}_r)$ be the set of r-step configuration sequences that starts from specific values of $\tilde{\mathbf{c}}_0$ and $\tilde{\mathbf{c}}_r$. The sum in Equation 17 could then be further rewritten as

$$\|\langle \phi_0 | T_r | \phi_r \rangle\|_\infty \leq \sum_{\widetilde{W}_r^{\mathbf{c}}(\tilde{\mathbf{c}}_0, \tilde{\mathbf{c}}_r)} \sum_{\tilde{\mathbf{c}}(\mathbf{c}_i) = \tilde{\mathbf{c}}_i} t_{\mathbf{c}}(\mathbf{c}_0, \cdots, \mathbf{c}_r) = \sum_{\mathcal{W}_r^{\mathbf{c}}(\mathbf{c}_0, \mathbf{c}_r)} \tilde{t}_{\mathbf{c}}(\tilde{\mathbf{c}}_0, \cdots, \tilde{\mathbf{c}}_r) \qquad (18)$$

where $\tilde{t}_{\mathbf{c}}(\tilde{\mathbf{c}}_0, \cdots, \tilde{\mathbf{c}}_r)$ sums over all sequences $(\mathbf{c}_0, \mathbf{c}_1, \cdots, \mathbf{c}_r)$ for which $\tilde{\mathbf{c}}(\mathbf{c}_i) = \tilde{\mathbf{c}}_i$. Figure 2 summarizes the relationship between the space of H eigenstates (restricted to the ancilla qubits), configurations and reduced configurations.

Let $\widetilde{\mathcal{C}}_-$ be the set of reduced configurations corresponding to basis states in $\mathcal{L}_-$. Then since for each ancilla register the low energy subspace is spanned by states with 0 or k qubits in $|1\rangle$ state, we have

$$\widetilde{\mathcal{C}}_- = \{i \in [m] | (\underbrace{k, k, \cdots, k}_{i \text{ times}}, 0, 0, \cdots, 0)\}. \qquad (19)$$

This property will become useful in our later discussion of Algorithm 2 in Section 4.2. Combining Lemma 1 and Equation 18 yields an upper bound to $\|T_r\|_2$ with *fixed* $\tilde{\mathbf{c}}_0 = (0, 0, \cdots, 0)$

as

$$\|T_r\|_2 \leq \sum_{\tilde{\mathbf{c}}_r \in \tilde{\mathcal{C}}_-} \sum_{\widetilde{\mathcal{W}}_r^{\mathbf{c}}(\tilde{\mathbf{c}}_0, \tilde{\mathbf{c}}_r)} \tilde{t}_{\mathbf{c}}(\tilde{\mathbf{c}}_0, \tilde{\mathbf{c}}_1, \cdots, \tilde{\mathbf{c}}_r) \tag{20}$$

with the function t defined in Equation 18. Equation 20 is the basis of our main algorithms for finding an upper bound to $\|T_r\|_2$ to be presented in Section 4.2. In order to illustrate the definitions and their properties introduced in this section, in Appendix 1 we present a detailed example showing the procedure for estimating $\|T_2\|_2$. The algorithms for general $\|T_r\|_2$ are discusssed in the subsequent Section 4.2.

4.2   *Algorithm for computing the upper bound*

An upper bound to the right hand side of Equation 20 can be efficiently evaluated and expressed using *monomial symmetric polynomials* in the λ_i coefficients [28]. A monomial symmetric polynomial $m_{\mathbf{b}}(x_1, x_2, \cdots, x_n) = \sum_{\pi} x_{\pi(1)}^{b_1} x_{\pi(2)}^{b_2} \cdots x_{\pi(k)}^{b_k}$ where $\mathbf{b} \in \mathbb{N}^k$ is the *partition* of the symmetric polynomial and $\pi : [n] \mapsto [k]$ is an arrangement of k elements among n elements. In order to address the combinatorics involved in summing over $\tilde{\mathbf{c}}$ sequences, we also need a $(k+1) \times (k+1)$ matrix M with element M_{ij} being the number of possible ways to cause a transition from an H eigenstate with energy E_i to an eigenstate with energy E_j by one application of the perturbation V. For the unperturbed Hamiltonian defined according to Equation 4, we have

$$M_{ij} = \begin{cases} i & \text{if } j = i - 1 \\ k - i & \text{if } j = i + 1 \\ 0 & \text{otherwise.} \end{cases} \tag{21}$$

With the above definition in place, here we present a simple method for evaluating the sum in Equation 18. Consider a fixed sequence of reduced configurations $\tilde{\mathbf{c}}_0, \tilde{\mathbf{c}}_1, \cdots, \tilde{\mathbf{c}}_r$ with $\tilde{\mathbf{c}}_0 = (0, 0, \cdots, 0)$. We compute $w = \tilde{t}_{\mathbf{c}}(\tilde{\mathbf{c}}_1, \tilde{\mathbf{c}}_2, \cdots, \tilde{\mathbf{c}}_r)$, the term in Equation 18 corresponding to $\tilde{\mathbf{c}}_0, \tilde{\mathbf{c}}_1, \cdots, \tilde{\mathbf{c}}_r$, by Algorithm 1, which can be considered as a procedure for computing the symmetric polynomial that is an upper bound to the right hand side of (22) for general r. The discussion in Appendix 1 may be used as an example for bounding the second order term[b].

In Algorithm 1, we assume (step 1) that $\tilde{\mathbf{c}}_0$ is the all-zero vector, as a consequence of the discussion that leads to Equation 15. Then in step 2 we check if the input sequence of reduced configurations would actually produce a non-zero contribution to T_r by examining the three criteria listed in Section 4.1 after Equation 9. Step 2a, 2b and 2c examines violation of each of the three criteria respectively. If any of the criteria is violated, return $w = 0$ since the input sequence does not contribute non-trivially to T_r. Step 3 introduces the data structure used for representing a monomial symmetric polynomial, which includes 1) a partition vector $\mathbf{b}$, 2) a reduced configuration $\tilde{\mathbf{c}}$ as given by the input sequence and 3) an injective partial function

---

[b] From (A.3) in Appendix 1 we can see that $M_{01} = 3$. The sequences in (A.3) give rise to the term whose norm is $3(\lambda_1^2 + \lambda_2^2) \cdot \frac{1}{z - E_1}$, which can be interpreted as a symmetric polynomial (recall that we have fixed $\tilde{\mathbf{c}}_0 = (0, 0)$)

$$\frac{1}{|z - E_1|} \cdot \Omega_1 \Omega_2 m_{(2)}(\lambda_1, \lambda_2) = \sum_{\substack{\tilde{\mathbf{c}}_1, \tilde{\mathbf{c}}_2 \text{ such that} \\ (\tilde{\mathbf{c}}_0, \tilde{\mathbf{c}}_1, \tilde{\mathbf{c}}_2) \in \widetilde{W}_2^{\mathbf{c}}}} t(\tilde{\mathbf{c}}_0, \tilde{\mathbf{c}}_1, \tilde{\mathbf{c}}_2) \tag{22}$$

with $\Omega_1 = M_{01} = 3$ and $\Omega_2 = M_{10} = 1$. Note that the right hand side of (22) matches with that of Equation 20 for $r = 2$.

Algorithm 1 : $w =\text{WALKBOUND}(\tilde{\mathbf{c}}_1, \tilde{\mathbf{c}}_2, \cdots, \tilde{\mathbf{c}}_r)$

---

1. Let $\tilde{\mathbf{c}}_0 = (\underbrace{0, 0, \cdots, 0})$.

2. Check if the following holds:

   (a) $\tilde{\mathbf{c}}_r \notin \widetilde{\mathcal{C}}_-$;
   (b) there is any pair of reduced configurations $\tilde{\mathbf{c}}_i = \tilde{\mathbf{c}}_{i+1}$.
   (c) any of the reduced configuration $\tilde{\mathbf{c}}_i$ with $i \in [r-1]$ satisfies $\mathbf{n}(\tilde{\mathbf{c}}_i) \notin \mathcal{N}_+$;

   If either of the conditions hold, $w = 0$ and return.

3. For each $i = 0, \cdots, r$, introduce a partition vector $\mathbf{b}_i$ of length at most i and a mapping $\mu_i : \tilde{\mathbf{c}}_i \mapsto \mathbf{b}_i$ that maps some elements of $\tilde{\mathbf{c}}_i$ to $\mathbf{b}_i$. We would like each element in $\mathbf{b}_i$ to have a unique pre-image in $\mathbf{c}_i$, because intuitively, the q-th element of $\mathbf{b}_i$, denoted as $b_{i,q}$, represents how many times the register $\mu^{-1}(b_{i,q})$ has been acted on by V during the sequence.

4. Start from $\mathbf{b}_0 = \emptyset$ and $\mu_0 = \emptyset$. We scan from $\tilde{\mathbf{c}}_1$ through $\tilde{\mathbf{c}}_r$ and update the μ and $\mathbf{b}$ assignments in the following way. Suppose we have already computed μ_i and $\mathbf{b}_i$. Then we find the element $\tilde{c}_{i+1,s}$ in $\tilde{\mathbf{c}}_{i+1}$ that differs from the corresponding element $\tilde{c}_{i,s}$ in $\tilde{\mathbf{c}}_i$.

   (a) If $\tilde{c}_{i,s}$ is not in the domain of μ_i (implying $\tilde{c}_{i,s} = 0$ since the s-th element of $\tilde{\mathbf{c}}$ has not been modified by the algorithm before), let $\mathbf{b}_{i+1} = \mathbf{b}_i \cup \{b_t\}$ with $b_t = 1$, $\tilde{c}_{i+1,s} = 1$ and $\mu_{i+1} = \mu_i \cup \{\tilde{c}_{i+1,s} \mapsto b_t\}$;
   (b) If $\tilde{c}_{i,s}$ is in the domain of μ_i, then first let $\mathbf{b}_{i+1} = \mathbf{b}_i$ and $\mu_{i+1} = \mu_i$ and then increment $\mu_{i+1}(\tilde{\mathbf{c}}_{i+1})$ by 1;
   (c) Compute and store $\Omega_i = M_{xy}$ with $x = \tilde{c}_{i,s}$ and $y = \tilde{c}_{i+1,s}$.

5. Return $w(\tilde{\mathbf{c}}_1, \tilde{\mathbf{c}}_2, \cdots, \tilde{\mathbf{c}}_{r-1}) = \left(\Pi_{i=1}^{r-1} |z - E(\tilde{\mathbf{c}}_i)|^{-1} \right) (\Pi_{i=1}^r \Omega_i) m_{\mathbf{b}_r}(\lambda_1, \lambda_2, \cdots, \lambda_m)$.

$\mu : \tilde{\mathbf{c}} \mapsto \mathbf{b}$. The partition $\mathbf{b}$ is needed for computing the symmetric polynomial $m_{\mathbf{b}}(\boldsymbol{\lambda})$, while $\tilde{\mathbf{c}}$ and μ are needed for guiding the computation of $\mathbf{b}$. As shown in step 4, depending on how the reduced configuration changes from the current i-th step to the $(i+1)$-st, the data structure for the new step is updated. Because at this point of the algorithm $\tilde{\mathbf{c}}_i$ and $\tilde{\mathbf{c}}_{i+1}$ differ by exactly one element (step 2 has excluded all invalid sequences of reduced configurations), what matters is then the position of the differing element, which we call $\tilde{c}_{i,s}$. There are two possibilities, each handled by step 4a and 4b. Recall that the value of each element in a reduced configuration stands for the energy level of an ancilla register. Making a transition from energy level $x = \tilde{c}_{i,s}$ to $y = \tilde{c}_{i+1,s}$ induces a combinatorial factor M_{xy} which is handled in step 4c.

The steps for computing μ_r and $\mathbf{b}_r$ takes $O(r)$ time. In the final step, evaluating the symmetric polynomial in w takes $O(m^r)$ time. Hence for a fixed r, the total time needed for computing the upper bound w for a *fixed* sequence of reduced configurations is polynomial in m. We outline the algorithm for computing the tight error bound in Algorithm 2. Recall the definition of $\boldsymbol{\lambda}$ at the beginning of Section 4.1.

Algorithm 2 : $\tau_r = \text{PERTURBBOUND}(r, k, \boldsymbol{\lambda}, M)$

---

1. If $r \geq k$, compute integers $p = \lfloor r/k \rfloor$ and $q = r - p$. Otherwise let $p = 0$, $q = 0$.

2. If q is odd, let $\gamma_r = 0$. Otherwise, compute

$$\gamma_r = \sum_{\substack{\tilde{\mathbf{c}}_1, \cdots, \tilde{\mathbf{c}}_{r-1} \text{ such that} \\ (\tilde{\mathbf{c}}_0, \cdots, \tilde{\mathbf{c}}_{r-1}, \tilde{\mathbf{c}}_0) \in \widetilde{\mathcal{W}}_r^{\mathbf{c}}}} \text{WALKBOUND}(\tilde{\mathbf{c}}_1, \cdots, \tilde{\mathbf{c}}_{r-1}, \tilde{\mathbf{c}}_0). \tag{23}$$

3. If $r \geq k$, then for each i from 1 to $\min\{p, m\}$

   (a) Let $\tilde{\mathbf{c}}_r = (\underbrace{k, k, \cdots, k}_{i \text{ times}}, \underbrace{0, 0, \cdots, 0}_{m - i \text{ times}})$;

   (b) Compute $\gamma_{i,r} = \sum_{\substack{\tilde{\mathbf{c}}_1, \cdots, \tilde{\mathbf{c}}_{r-1} \text{ such that} \\ (\tilde{\mathbf{c}}_0, \cdots, \tilde{\mathbf{c}}_{r-1}, \tilde{\mathbf{c}}_r) \in \widetilde{\mathcal{W}}_r^{\mathbf{c}}}} \text{WALKBOUND}(\tilde{\mathbf{c}}_1, \cdots, \tilde{\mathbf{c}}_{r-1}, \tilde{\mathbf{c}}_r).$

   Otherwise let $\gamma_{i,r} = 0$ for all i.

4. Return $\tau_r = \gamma_r + \sum_{i=1}^{\min\{p,m\}} \gamma_{i,r}$.

Algorithm 2 essentially computes the right hand side of Equation 20. Step 1 computes parameters p and q that are relevant to the the general structure of $\mathcal{W}_r$ introduced in the beginning of Section 4.1. Recall the three criteria for sequences $(\phi_0, \phi_1, \cdots, \phi_r)$ in $\mathcal{W}_r$: ϕ_0 and ϕ_r must be of the form $s_1 s_2 \cdots s_m$ with each $s_i \in \{0^k, 1^k\}$ while $\phi_1, \cdots, \phi_{r-1}$ must not be of this form and ϕ_i, ϕ_{i+1} must differ by one and only one bit for $i = 0, \cdots, r - 1$. For $r < k$, clearly $\mathcal{W}_r$ could only contain sequences where $\phi_0 = \phi_r$, resulting in T_r being proportional to the identity operator in $\mathcal{L}_-$ with the proportional constant being γ_r (Equation 6). For $r \geq k$, it is then possible that $\phi_r = s_1 s_2 \cdots s_m$ with at most $p = \lfloor r/k \rfloor$ substrings $s_i = 1^k$. Therefore

one needs to also take sequences with these ϕ_r possibilities into account. Step 2 computes the magnitude of the term in T_r that is proportional to identity in the low energy subspace $\mathcal{L}_-$, by summing over all sequences of reduced configurations that starts from $\tilde{\mathbf{c}}_0$ and ends at $\tilde{\mathbf{c}}_0$. For $r \leq k$, the γ_r computed here is precisely the leading coeffcients γ_r in Equation 6. Looping over all viable sequences $(\tilde{\mathbf{c}}_0, \tilde{\mathbf{c}}_1, \cdots, \tilde{\mathbf{c}}_{r-1}, \tilde{\mathbf{c}}_0)$ induces $O(m^{k(r-1)})$ in the computational cost. Step 3 completes the remainder of the outer summation in Equation 20 over the space $\widetilde{\mathcal{C}}_-$ for $\tilde{\mathbf{c}}_r$. In the case where $r < k$ this step is entirely skipped and there will be no $\gamma_{i,r}$ values computed. There are in total at most m iterations in the step and each iteration sum over at most $O(m^{kr})$ sequences $(\tilde{\mathbf{c}}_0, \tilde{\mathbf{c}}_1, \cdots, \tilde{\mathbf{c}}_r)$. Hence the total runtime of Algorithm 2 scales as $O(m^{kr+1})$, which is polynomial in the number m of ancilla registers.

4.3 Sharpness of the upper bound

Algorithm 2 is useful for computing the exact value of γ_r in T_r with $r < k$ (Equation 6). Here we show that for the special case where the terms $H_{\text{eff},i}$ pairwise commute, Algorithm 2 also allows one to efficiently compute the exact value of $\|T_r\|_2$ for any $r \geq k$. This is a stronger claim than the one implied by Equation 20 and we state it precisely in the following Theorem 2.

Theorem 2 *Given the k-body target Hamiltonian $H_{targ} = \sum_{i=1}^{m} c_i H_{targ,i}$ as defined in Section 2, the gadget Hamiltonian $\tilde{H} = H + V$ as defined in Equations 4 and 5, and the self energy expansion $\Sigma_-(z)$ shown in Equation 6. If for any $i, j \in \{1, \cdots, m\}$, $[H_{targ,i}, H_{targ,j}] = 0$, then for any $r \geq 2$ and $k \geq 3$, we have*

$$\|T_r\|_2 = \text{PERTURBBOUND}(r, k, \boldsymbol{\lambda}, M) \tag{24}$$

where $\boldsymbol{\lambda}$ is defined in Section 4.1 and M is defined in Equation 21.

Before proving Theorem 2, we would like to first establish a few properties of the sets $\mathcal{W}_r$, $\mathcal{W}_r^{\mathbf{c}}$, and $\widetilde{W}_r^{\mathbf{c}}$, which are introduced in Section 4.1. Recall that $\mathcal{W}_r$ is a collection of sequences $(\phi_0, \phi_1, \cdots, \phi_r)$ with each ϕ_i being a km-bit string which we write as $s_1 s_2 \cdots s_m$, $s_j \in \{0, 1\}^k$. Let ζ be a permutation that only involves permuting bits inside the same k-bit substrings s_j. In other words, for any bit $b \in \phi = s_1 s_2 \cdots s_m$, let j be such that $b \in s_j$, then $\zeta(b) \in s_j$ always holds. We call such ζ a *local permutation*. For any sequence $W = (\phi_0, \phi_1, \cdots, \phi_r)$, we use the notation $\zeta(W) = (\zeta(\phi_0), \zeta(\phi_1), \cdots, \zeta(\phi_r))$ to mean a sequence produced by applying the permutation ζ onto the string at every intermediate step. We further say that for two sequences A and B in $\mathcal{W}_r$, $A \sim B$ if there is a local permutation ζ such that $A = \zeta(B)$. Clearly the relation $\sim$ is symmetric, reflexive and transitive. Hence $\sim$ is an equivalence relation that partitions $\mathcal{W}_r$ into equivalence classes (Figure 3). Because local permutations do not alter the Hamming weight of any k-bit substrings s_i, each equivalence class in $\mathcal{W}_r$ corresponds to an element in $\mathcal{W}_r^{\mathbf{c}}$ since a configuration $\mathbf{c}$ is constructed based on Hamming weights of substrings in a state $|\phi\rangle$ (Figure 2). In the set $\mathcal{W}_r^{\mathbf{c}}$ we could also define an equivalence relation $\sim^{\mathbf{c}}$ based on permutation over the ancilla registers *i.e.* the elements of the configuration vector. For A, B in $\mathcal{W}_r^{\mathbf{c}}$, we say $A \sim^{\mathbf{c}} B$ if there is a permutation π over ancilla registers such that $A = \pi(B)$. Here the definition of π acting on a sequence of configurations is analogous to ζ on a sequence of ancilla states and we omit the details. Since by definition reduced configurations are obtained from sorting the elements of configurations, each equivalence class in $\mathcal{W}_r^{\mathbf{c}}$ naturally corresponds to a sequence of reduced configuration in $\widetilde{\mathcal{W}}_r^{\mathbf{c}}$.

Proof of Theorem 2. Because we could express any $\phi \in \mathcal{L}_-$ as $|\phi\rangle = S|0^{mk}\rangle$ for some S with the form in Equation 12, we could write T_r as (cf. Equation 14)

$$T_r = \sum_{S_1,S_2} \langle 0^{km}|S_1 T_r S_2|0^{km}\rangle S_1 |0^{km}\rangle\langle 0^{km}|S_2 \tag{25}$$

where the summation is over any pair of operators S_1, S_2 of the form in Equation 12. We could further split T_r as a sum of diagonal and off-diagonal components:

$$
\begin{aligned}
T_r &= \sum_S \langle 0^{km}|S T_r S|0^{km}\rangle + \sum_{S_1}\sum_{S_2 \neq S_1} \langle 0^{km}|T_r S_1 S_2|0^{km}\rangle S_1 |0^{km}\rangle\langle 0^{km}|S_2 \\
&= \sum_S \langle 0^{km}|T_r|0^{km}\rangle S|0^{km}\rangle\langle 0^{km}|S + \sum_{S_1}\sum_{S \neq I} \langle 0^{km}|T_r S|0^{km}\rangle S_1|0^{km}\rangle\langle 0^{km}|S_1 S \\
&= \langle 0^{km}|T_r|0^{km}\rangle \sum_S S|0^{km}\rangle\langle 0^{km}|S + \sum_{S \neq I} \langle 0^{km}|T_r S|0^{km}\rangle \sum_{S_1} S_1(|0^{km}\rangle\langle 0^{km}|S)S_1 \\
&= \sum_S O_{S,r} \otimes \Pi_S
\end{aligned}
$$
$$\tag{26}$$

where going from the first line to the second we have used Equation 13 on the first term and applied the substitution $S_2 = S_1 S$ to the second term. Going from the second line to the third is merely a relocation of the summation so that the form of the expression can be more easily recognized as the last line with

$$O_{S,r} = \langle 0^{km}|T_r S|0^{km}\rangle \quad \text{and} \quad \Pi_S = \sum_{S'} S'(|0^{km}\rangle\langle 0^{km}|S)S'. \tag{27}$$

Note that the projectors Π_S have unit norms and they are orthogonal, namely $\|\Pi_{S_1}\Pi_{S_2}\| = \delta_{S_1,S_2}$. Since each $H_{\mathrm{targ},i}$ is a tensor product of Pauli operators and we assume that the set of $H_{\mathrm{targ},i}$ terms pairwise commute, we have

$$\|T_r\|_2 = \sum_S \|O_{S,r}\|_2 = \sum_S \|O_{S,r}\|_\infty. \tag{28}$$

Because the operator S flips all of the qubits in a certain subset $F \subseteq \{1, 2, \cdots, m\}$ of ancilla registers, the operator $O_{S,r}$ is proportional to $\sum_{i \in F} H_{\mathrm{targ},i}$ (see Equation 3 for the definition of $H_{\mathrm{targ},i}$). Therefore the problem of evaluating $\|T_r\|_2$ becomes the problem of finding the coefficients for all of the $O_{S,r}$ operators. These coefficients are precisely given by Algorithm 2. We show that in fact the quantities γ_r computed at step 2 and $\gamma_{i,r}$ computed at step 3b satisfy

$$
\begin{aligned}
\gamma_r &= \|\langle 0^{km}|T_r|0^{km}\rangle\|_\infty \\
\gamma_{i,r} &= \sum_{\substack{S \text{ acting} \\ \text{on } i \text{ regs}}} \|\langle 0^{km}|T_r S|0^{km}\rangle\|_\infty.
\end{aligned}
\tag{29}
$$

Here in the expression for $\gamma_{i,r}$ the summation is over all S operators of the form in Equation 13 that acts non-trivially on i ancilla registers. The collection of all states of the form $S|0^{km}\rangle$ acting non-trivially on i registers is then the set of all states $|\phi\rangle$, $\phi \in \{0,1\}^{km}$ such that its reduced configuration is $\tilde{\mathbf{c}} = (\underbrace{k, \cdots, k}_{i \text{ times}}, 0, \cdots, 0)$.
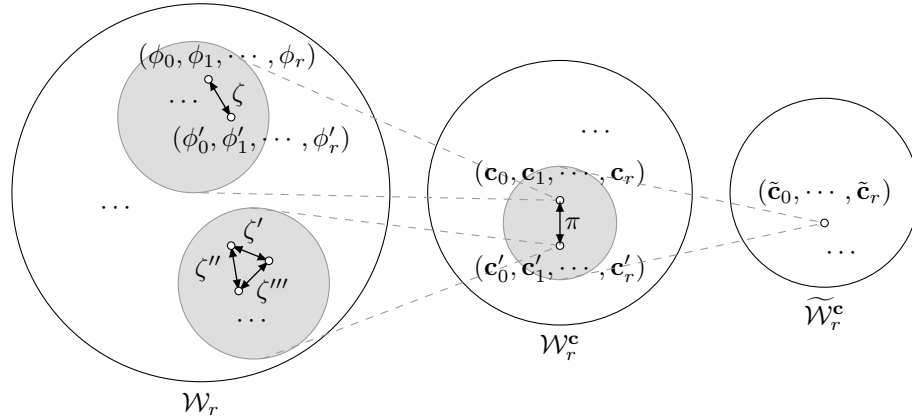
Fig. 3. The hierarchy of equivalence classes that relates the sets $\mathcal{W}_r$, $\mathcal{W}_r^{\mathbf{c}}$, and $\widetilde{\mathcal{W}}_r^{\mathbf{c}}$. Each light shaded circle represents an equivalence class inside the set. For $\mathcal{W}_r$, each node represents a sequence and sequences in the same equivalence class are related by a "local permutation" (defined formally in Section 4.3) ζ that only permutes qubits in the same ancilla register. Each equivalence class in $\mathcal{W}_r$ then maps to a sequence of configurations in $\mathcal{W}_r^{\mathbf{c}}$, as shown with the dashed lines, and each sequence of configurations in $\mathcal{W}_r^{\mathbf{c}}$ also belongs to an equivalence class where the sequences are related by permutation π of the elements of the configuration *i.e.* ancilla registers. Each equivalence class in $\mathcal{W}_r^{\mathbf{c}}$ then corresponds to an element in $\widetilde{\mathcal{W}}_r^{\mathbf{c}}$.

We show Equation 29 by taking advantage of the hierarchical structure of equivalence classes in the sets $\mathcal{W}_r$, $\mathcal{W}_r^{\mathbf{c}}$, and $\widetilde{W}_r^{\mathbf{c}}$ (Figure 3). Observing Equation 9 and Equation 16 we see that $t_\phi(\phi_0, \phi_1, \cdots, \phi_r)$ is in fact the norm of the $|\phi_0\rangle\langle\phi_r|$ block of T_r. For a specific $|\phi_r\rangle = S|0^{km}\rangle$ for some S, we could evaluate $\|O_{S,r}\|_2$ by summing over the weight t_ϕ of all sequences in $\mathcal{W}_r(0^{km}, \phi_r)$. We partition the set $\mathcal{W}_r(0^{km}, \phi_r)$ into equivalence classes as discussed before and each equivalence class is associated with an element in $\mathcal{W}_r^{\mathbf{c}}(\mathbf{c}_0, \mathbf{c}_r)$ with $\mathbf{c}_0$, $\mathbf{c}_r$ being the configurations of $|0^{mk}\rangle$ and $S|0^{mk}\rangle$ respectively. We could evaluate the weight $t_{\mathbf{c}}$ of a configuration sequence $(\mathbf{c}_0, \mathbf{c}_1, \cdots, \mathbf{c}_r) \in \mathcal{W}_r^{\mathbf{c}}$ as the sum of the weights of all the elements in the equivalent class in $\mathcal{W}_r$ associated with it. Specifically, starting with any sequence $(\phi_0, \cdots, \phi_r)$ in the equivalence class, we could calculate the weight $t_{\mathbf{c}}$ by summing over all the local permutations ζ that produces the elements in the class:

$$t_{\mathbf{c}}(\mathbf{c}_0, \mathbf{c}_1, \cdots, \mathbf{c}_r) = \sum_{\zeta} t_\phi(\zeta(\phi_0), \zeta(\phi_1), \cdots, \zeta(\phi_r)). \tag{30}$$

At each step of a sequence in $\mathcal{W}_r$ going from ϕ_{i-1} to ϕ_i, suppose the number of $|1\rangle$ ancilla qubits in the register p_i changes from x_i to y_i. To evaluate Equation 30, we need to sum over all possible ways in which a k-qubit state with Hamming weight x_i can make a transition to a state with Hamming weight y_i through the action of the perturbation V. Because local permutations do not change the weight of the sequence $(\phi_0, \cdots, \phi_r)$, the weight $t_{\mathbf{c}}(\mathbf{c}_0, \cdots, \mathbf{c}_r)$ differs from $t_\phi(\phi_0, \cdots, \phi_r)$ by a multiplicative factor. With the specific construction of V in Equation 5, such multiplicative factor can be calculated using M_{ij} in Equation 21. Let $\Omega_i = M_{x_i y_i}$, which is the number of possible ways for an ancilla register to go from a state

with x_i qubits in $|1\rangle$ to one with y_i qubits in $|1\rangle$. Then we have

$$t_{\mathbf{c}}(\mathbf{c}_0, \mathbf{c}_1, \cdots, \mathbf{c}_r) = \left(\prod_{i=1}^{r} \Omega_i \right) t_\phi(\phi_0, \phi_1, \cdots, \phi_r) \qquad (31)$$

with $(\phi_0, \phi_1, \cdots, \phi_r)$ being any sequence with the configuration of ϕ_i being $\mathbf{c}_i$. Similar to Equation 30, we could also evaluate the weight of a sequence of reduced configuration by summing over all permutations of m registers:

$$
\begin{aligned}
\tilde{t}_{\mathbf{c}}(\tilde{\mathbf{c}}_0, \tilde{\mathbf{c}}_1, \cdots, \tilde{\mathbf{c}}_r) &= \sum_{\pi:[m]\mapsto[m]} t_{\mathbf{c}}(\pi(\mathbf{c}_0), \pi(\mathbf{c}_1), \cdots, \pi(\mathbf{c}_r)) \\
&= \prod_{i=1}^{r-1} \frac{1}{|z - E(\tilde{\mathbf{c}}_i)|} \prod_{i=1}^{r} \Omega_i \underbrace{\sum_{\pi:[m]\mapsto[m]} \lambda_{\pi(p_1)} \lambda_{\pi(p_2)} \cdots \lambda_{\pi(p_r)}}_{(*)}.
\end{aligned}
\qquad (32)
$$

Here the term $(*)$ is essentially a monomial symmstric polynomial over the variables $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \cdots, \lambda_m)$. We can rewrite it as $m_{\mathbf{b}_r}(\boldsymbol{\lambda})$ with $\mathbf{b}_r$ being the partition of the symmetric polynomial that keeps track of "how many registers have been acted on by how many times". For example, if there are three registers that are acted on (i.e. have one or more bits flipped in them) once and one register acted on twice, in which case the order of perturbation theory is $r = 3 \times 1 + 1 \times 2 = 5$, then $\mathbf{b}_r = (2, 1, 1, 1)$. In general one could compute $\mathbf{b}_r$ for a reduced configuration sequence in $\widetilde{\mathcal{W}}_r^{\mathbf{c}}$.

From the arguments so far, it should be clear that Algorithm 1 computes $\tilde{t}_{\mathbf{c}}$ correctly. Because for any S of the form in Equation 12, $S|0^{km}\rangle$ is the only state with its configuration and reduced configuration, summing the weights of all sequences $(0^{km}, \phi_1, \cdots, \phi_r)$ with $|\phi_r\rangle = S|0^{km}\rangle$ in $\mathcal{W}_r$ is equivalent to summing the weights of all sequences of reduced configurations with $\tilde{\mathbf{c}} = (0, 0, \cdots, 0)$ and $\tilde{\mathbf{c}}_r = (k, \cdots, k, 0, \cdots, 0)$ with the number of elements equal to k being the number of registers that S acts on. Therefore Equation 29 holds and the main statement of the theorem is proven. $\qquad\square$

5   Potential improvement using Schrieffer-Wolff transformation

Theorem 2 presented in Section 4.3 shows the sharpness of the bounds provided by our algorithm for difference in norm $\|\Sigma_-(z) - H_{\text{eff}}\|_2$ between the self-energy $\Sigma_-(z)$ and the effective Hamiltonian H_{eff} in Equation 6. However, the quantity $\|\Sigma_-(z) - H_{\text{eff}}\|_2$ is itself an upper bound to the actual spectral difference between the gadget and target Hamiltonian (see Theorem 1 as well as numerics in Section 6) because $\|\Sigma_-(z) - H_{\text{eff}}\|_2 \le \epsilon$ is only a *sufficient* condition that guatantees that the spectral difference is small, namely $\max_j |\lambda_j(\tilde{H}) - \lambda_j(H_{\text{targ}})| \le \epsilon$ with $\lambda_j(H)$ being the j-th lowest eigenvalue of H. In practice one typically is more concerned about the spectral difference between the target and gadget Hamiltonian than $\|\Sigma_-(z) - H_{\text{eff}}\|_2$. Therefore to obtain a tighter upper bound to the actual spectral error than the norm difference $\|\Sigma_-(z) - H_{\text{eff}}\|_2$ based on Feynman-Dyson (FD) series, we need to adopt a different formalism of perturbation theory. Here we consider using the Schrieffer-Wolff (SW) transformation. As we will prove in this section, the SW series generates all the terms in the FD series but also includes more terms that are beyond FD. In Section 6 we

numerically show an improved bound for the spectral error based on the SW transformation over $\|\Sigma_-(z) - H_{\text{eff}}\|_2$.

The Schrieffer-Wolff transformation is a formalism of degenerate perturbation theory where the low energy effective Hamiltonian H_{eff} is obtained from the perturbed Hamiltonian by a unitary transformation that makes the perturbed Hamiltonian block diagonal with respect to low and high energy subspaces [31]. Using the same setting as Section 2, we consider a perturbed Hamiltonian $\tilde{H} = H + V$ which is a sum of some unperturbed Hamiltonian H that is diagonal in the basis we are assuming and a perturbation $V = V_d + V_{od}$ that contains both diagonal V_d and off-diagonal V_{od} components. The basic idea of SW transformation is to find an anti-Hermitian operator R such that $e^R(H + V)e^{-R}$ is block diagonal with respect to the high energy subspace $\mathcal{L}_+$ and the low energy subspace $\mathcal{L}_-$. The effective low energy Hamiltonian is then $H_{\text{eff}} = \Pi_- e^R(H + V)e^{-R}\Pi_-$ with Π_- being the projector to $\mathcal{L}_-$.

We denote the adjoint operation of an operator Y on X as $\hat{Y}(X) = [Y, X]$. Let $\mathcal{O}$ be a super operator that extracts the off-diagonal component of an operator. For example $V_{od} = \mathcal{O}(V)$. Let $\mathcal{K}$ be a super operator such that

$$\mathcal{K}(X) = \sum_{i,j} \frac{\langle i|\mathcal{O}(X)|j\rangle}{E_i - E_j}|i\rangle\langle j| \tag{33}$$

where the $|i\rangle$, $|j\rangle$ states are the eigenstates of H and the summation is over any i, j such that either $|i\rangle \in \mathcal{L}_-$, $|j\rangle \in \mathcal{L}_+$, or $|i\rangle \in \mathcal{L}_+$, $|j\rangle \in \mathcal{L}_-$.

The anti-Hermitian operator R admits an expansion $R = \sum_{n=1}^{\infty} R_n$. To ensure that the transformed Hamiltonian $e^{\hat{R}}(H + V) = e^R(H + V)e^{-R}$ is block diagonal, the R_n terms are given by[c]

$$\begin{aligned} R_1 &= \mathcal{K}(V_{od}) \\ R_2 &= \mathcal{K}(\hat{R}_1 V_d) \\ R_n &= -\mathcal{K}\hat{V}_d(R_{n-1}) + \sum_{j\geq 1} a_{2j}\mathcal{K}\hat{R}^{2j}(V_{od})_{n-1} \end{aligned} \tag{34}$$

where $a_m = \frac{2^m}{m!}B_m$, B_m being the m-th Bernoulli number and

$$\hat{R}^k(V_{od})_m = \sum_{\substack{n_1+\cdots+n_k=m \\ n_1,\cdots,n_k\geq 1}} \hat{R}_{n_1}\hat{R}_{n_2}\cdots\hat{R}_{n_k}(V_{od}). \tag{35}$$

The effective Hamiltonian $H_{\text{eff,SW}}$ is then given by

$$H_{\text{eff,SW}} = H_- + V_- + \sum_{r=2}^{\infty} \underbrace{\sum_{j\geq 1} b_{2j-1}\Pi_-\hat{R}^{2j-1}(V_{od})_{r-1}\Pi_-}_{H_{\text{eff},r}} \tag{36}$$

where the coefficients $b_n = \frac{2(2^{2n}-1)}{(2n)!}B_{2n}$. Note that the summation in $H_{\text{eff},r}$ over $j \geq 1$ is not infinite, because from Equation 35 we see that $\hat{R}^k(V_{od})_m = 0$ if $k > m$. From the definition of $H_{\text{eff},r}$ it is clear that for any r, the r-th order effective Hamiltonian $H_{\text{eff},r}$ must contain a term of the form

$$-b_1\Pi_-\hat{V}_{od}(-\mathcal{K}\hat{V}_d)^{r-2}R_1\Pi_-. \tag{37}$$

---

[c]In [31] the authors consider a setting where $\tilde{H} = H + \varepsilon V$ and R is a Taylor expansion $R = \sum_{n=1}^{\infty} R_n\varepsilon^n$. Here we absorb the ε parameter into V and replace ε with 1.

There are of course other terms appearing at any order r and in [31] the authors have created an elegant diagrammatic technique for enumerating the terms. However, here we focus on the terms of the form in (37) (which correspond to tree diagrams that are simply a linear chain of nodes) and show that at any order n, in some sense (37) is equivalent to the entire r-th order term T_r in the FD series (Equations 1 and 2). We state it precisely in the following theorem.

Theorem 3 *If H has a unique eigenvalue E_0 that is below the cutoff between low and high energy subspaces $\mathcal{L}_-$ and $\mathcal{L}_+$, then for any $r \geq 2$*

$$T_r(E_0) = -b_1 \Pi_- \hat{V}_{od}(-\mathcal{K}\hat{V}_d)^{r-2} R_1 \Pi_-. \tag{38}$$

Here we write T_r explicitly as a function of z, namely $T_r(z) = V_{-+}(G_+(z)V_+)^{r-2}G_+V_{+-}$.

For simplicity, we denote the eigenstates of H as $|i\rangle$ with $H|i\rangle = E_i|i\rangle$. To prove the statement we first show that for any $r \geq 1$,

$$
(\mathcal{K}\hat{V}_d)^r R_1 = \sum_{i \in \mathcal{L}_-} \sum_{j_1 \in \mathcal{L}_+} \cdots \sum_{j_{r+1} \in \mathcal{L}_+} \langle i|V|j_1\rangle \frac{1}{E_i - E_{j_1}} \langle j_1|V|j_2\rangle \cdots
$$
$$
\cdots \langle j_r|V|j_{r+1}\rangle \left(\frac{1}{E_{j_{r+1}} - E_i} |j_{r+1}\rangle\langle i| + \frac{1}{E_i - E_{j_{r+1}}} |i\rangle\langle j_{r+1}| \right). \tag{39}
$$

We prove (39) inductively on r. The base case is $r = 1$, By straightforward calculation

$$
\mathcal{K}\hat{V}_d R_1 = \sum_{i \in \mathcal{L}_-} \sum_{j \in \mathcal{L}_+} \sum_{k \in \mathcal{L}_+} \langle i|V|j\rangle \frac{1}{E_i - E_j} \langle j|V|k\rangle \left(\frac{1}{E_k - E_i} |k\rangle\langle i| + \frac{1}{E_i - E_k} |i\rangle\langle k| \right). \tag{40}
$$

The case for general k can be proved by similar calculations using the definitions of $\mathcal{K}$ and V_d.

Proof of Theorem 3. Rewriting the projected operators V_{-+}, G_+ etc into a summation over $|i\rangle\langle j|$ blocks, we have for example

$$
V_{-+} = \sum_{i \in \mathcal{L}_-} \sum_{j \in \mathcal{L}_+} \langle i|V|j\rangle |i\rangle\langle j|, \qquad G_+(z) = \sum_{i \in \mathcal{L}_+} \frac{1}{z - E_i} |i\rangle\langle i| \tag{41}
$$

and similar for V_+ and V_{+-}. Hence we could rewrite $T_r(z)$ as

$$
V_{-+}(G_+(z)V_+)^{r-2}G_+(z)V_{+-} = \sum_{i \in \mathcal{L}_-} \sum_{j_1 \in \mathcal{L}_+} \cdots \sum_{j_{r-1} \in \mathcal{L}_+} \sum_{\ell \in \mathcal{L}_-} \langle i|V|j_1\rangle \frac{1}{z - E_{j_1}} \langle j_1|V|j_2\rangle \cdots
$$
$$
\cdots \frac{1}{z - E_{j_{r-1}}} \langle j_{r-1}|V|\ell\rangle |i\rangle\langle\ell|. \tag{42}
$$

Using Equation 39 and $\Pi_- = \sum_{i \in \mathcal{L}_-} |i\rangle\langle i|$, we have

$$
-b_1 \Pi_- \hat{V}_{od}(-\mathcal{K}\hat{V}_d)^{r-2} R_1 \Pi_- = \frac{1}{2} \sum_{i \in \mathcal{L}_-} \sum_{j_1 \in \mathcal{L}_+} \cdots \sum_{j_{r-1} \in \mathcal{L}_+} \sum_{\ell \in \mathcal{L}_-} \langle i|V|j_1\rangle \frac{1}{E_i - E_{j_1}} \langle j_1|V|j_2\rangle \cdots
$$
$$
\cdots \frac{1}{E_i - E_{j_{r-1}}} (|\ell\rangle\langle i| + |i\rangle\langle\ell|). \tag{43}
$$

Comparing Equations 43 with 42 and the main equation (38) follows.   □

6 Numerical example

6.1 PERTURBBOUND *vs. Simple upper bound*

Here we compare the tightness of bounds obtained by PERTURBBOUND and simple upper bounds (from the right hand side of Equation 7). Consider applying the gadget construction in Section 2 on the 3-body target Hamiltonian is $H_{\text{eff}} = \alpha_1 X_1 X_2 X_3 + \alpha_2 X_2 Y_4 Z_5$ where α_1 and α_2 are real coefficients (Figure 4b). The resulting gadget Hamiltonian is described in Figure 4a, which can be expressed in form of the general setting $\tilde{H} = H + V$. Here the unperturbed Hamiltonian H and perturbation V are defined as

$$
\begin{aligned}
H &= H^{(1)} + H^{(2)}, & H^{(1)} &= \frac{\Delta}{4}(3\mathbf{I} - Z_{u_1}Z_{u_2} - Z_{u_2}Z_{u_3} - Z_{u_1}Z_{u_3}) \\
& & H^{(2)} &= \frac{\Delta}{4}(3\mathbf{I} - Z_{v_1}Z_{v_2} - Z_{v_2}Z_{v_3} - Z_{v_1}Z_{v_3}) \\
V &= V^{(1)} + V^{(2)}, & V^{(1)} &= \mu_1(X_1 X_{u_1} + X_2 X_{u_2} + X_3 X_{u_3}) \\
& & V^{(2)} &= \mu_2(Y_4 X_{v_1} + X_2 X_{v_2} + Z_5 X_{v_3})
\end{aligned}
\tag{44}
$$

where spins with u_i and v_i labels belong to the two unperturbed subsystems. Here we let Δ be orders of magnitude larger than μ_1 and μ_2 and keep the coefficients μ_1 and μ_2 as

$$
\mu_1 = \left(\frac{\alpha_1 \Delta^2}{6}\right)^{1/3}, \qquad \mu_2 = \left(\frac{\alpha_2 \Delta^2}{6}\right)^{1/3}
\tag{45}
$$

where α_1 and α_2 are parameters related to the low energy effective Hamiltonian (see Equation 49). In Figure 4c we explicitly partition the Hamiltonian in terms of H and V.

The low-energy subspace of the total Hamiltonian $\tilde{H}$ is then $\mathcal{L}_- = \mathcal{L}_-^{(1)} \otimes \mathcal{L}_-^{(2)}$. Inspecting the expressions $H^{(1)}$ and $H^{(2)}$ gives the low energy subspaces for each subsystem: $\mathcal{L}_-^{(1)} = \text{span}\{|000\rangle_{u_1 u_2 u_3}, |111\rangle_{u_1 u_2 u_3}\}$ and $\mathcal{L}_-^{(2)} = \text{span}\{|000\rangle_{v_1 v_2 v_3}, |111\rangle_{v_1 v_2 v_3}\}$. For each subsystem $i \in \{1, 2\}$, the subspaces of $H^{(i)}$ and their corresponding energies are

$$
\begin{aligned}
\mathcal{P}_0 &= \text{span}\{|000\rangle\}, & E_0 &= 0 \\
\mathcal{P}_1 &= \text{span}\{|001\rangle, |010\rangle, |100\rangle\}, & E_1 &= \Delta \\
\mathcal{P}_2 &= \text{span}\{|011\rangle, |101\rangle, |110\rangle\}, & E_2 &= \Delta \\
\mathcal{P}_3 &= \text{span}\{|111\rangle\}, & E_3 &= 0.
\end{aligned}
\tag{46}
$$

In Figure 4d we show the spectrum of each subsystem. The matrix M defined in Equation 21 is also involved in the computation of the upper bound to $\|T_r\|$. We could interpret M from Figure 4d. One could regard M_{ij} as the maximum, over all eigenstates of H in $\mathcal{P}_i$, number of possible transitions from a particular $|u\rangle \in \mathcal{P}_i$ to an eigenstate in $\mathcal{P}_j$. Precisely,

$$
M_{ij} = \max_{|u\rangle \in \mathcal{P}_i} Card\{|v\rangle \in \mathcal{P}_j | \|\langle v|V|u\rangle\| \neq 0\}
\tag{47}
$$

where $Card\{\cdot\}$ stands for cardinality (number of distinct elements) of a set. We could then determine that

$$
M = \begin{array}{c@{}c}
 & \begin{array}{cccc} \mathcal{P}_0 & \mathcal{P}_1 & \mathcal{P}_2 & \mathcal{P}_3 \end{array} \\
\begin{array}{c} \mathcal{P}_0 \\ \mathcal{P}_1 \\ \mathcal{P}_2 \\ \mathcal{P}_3 \end{array} &
\left[\begin{array}{cccc}
 & 3 & & \\
1 & & 2 & \\
 & 2 & & 1 \\
 & & 3 &
\end{array}\right]
\end{array}
\tag{48}
$$

where the row and column indices start from 0 because the subspaces $\mathcal{P}_0$, $\mathcal{P}_1$, $\cdots$, have indices that start from 0.

From Figure 4a and 4c we can see that the unperturbed system H essentially consists of two identical 4-level systems with energy levels E_0, E_1, E_2 and E_3. This gives rise to in total 9 possible energy combinations.

With the matrix M worked out as in Equation 48, we could use the algorithm WALK-BOUND in Section 4.2 to find a tight upper bound for $\|T_r\|_\infty$ at any order r. After a certain order p, when the upper bound becomes less than the tolerance 10^{-8}, we use Equation 7 to bound the terms from $p+1$ to infinity.

Using the perturbation series in Equation (1) we could show that if we truncate the series at the 3rd order, namely $\Sigma_-(z) = H_{\text{eff}} + T_4 + T_5 + \cdots$, we have the effective 3-body Hamiltonian

$$H_{\text{eff}} = \alpha_1 X_1 X_2 X_3 + \alpha_2 X_2 Y_4 Z_5 + \gamma I \tag{49}$$

with γ being the magnitude of the spectral shift. Here we let $\alpha_1 = 0.1$ and $\alpha_2 = 0.2$. Then the entire Hamiltonian $\tilde{H} = H + V$ in Equation 44 is only dependent on a free parameter Δ. In order to test our algorithm for bounding perturbative terms, we treat terms from 4th order onward as errors in the perturbation series. This amounts to estimating $\|\Sigma_-(z) - H_{\text{eff}}\|_2$. We could compute this value by explicitly computing $\Sigma_-(z)$ by its definition $zI - (\tilde{G}_-(z))^{-1}$ and then evaluating $\|\Sigma_-(z) - H_{\text{eff}}\|_2$. This method is inefficient since it requires inverting an exponentially large matrix with respect to system size, but yields an accurate estimation for the error $\|\Sigma_-(z) - H_{\text{eff}}\|_2$. We will use it as a benchmark for comparison with the upper bound computed by the new algorithm developed here. As shown Figure 5, the upper bounds computed by PERTURBBOUND are tight with respect to the exact calculation. For the purpose of comparison we also compute the error bound due to triangle inequality (see Equation 7). We explicitly computed $\|V\|_2$ (while in practice one may use some upper bound for $\|V\|_2$ which could loosen the bound further but here for comparison we use the exact value) and bounded $\|G_+\|_2$ from above by $1/E_1$. Hence the simple bound of error terms from a certain order to infinity based on Equation 7 becomes $\sum_{r=4}^\infty \|V\|_2^r / E_1^{r-1} = \|V\|_2^4 / (E_1^2(E_1 - \|V\|_2))$. When implementing our algorithm for the numerical example concerned in this section, we compute $\tau_r = \text{PERTURBBOUND}(r, \boldsymbol{\lambda}, M)$ for r from 4 to a value p such that $\tau_p \leq 10^{-20}$. Then we resort to Equation 7 for computing an upper bound to $\|T_{p+1} + T_{p+2} + \cdots\|_2$.

The ultimate purpose for finding tight error bound in the perturbation theory is to find lower assignments to Δ while maintaining the spectral error between the target and the gadget Hamiltonian within ϵ. As mentioned in Section 3, with an algorithm for computing an upper bound to the spectral error we could find the optimal Δ assignment based on this algorithm by using binary search to find a Δ such that the error bound is ϵ. In Figure 6 we show the result of implementing such binary search for three means for estimating the spectral error: 1) crude upper bound based on geometric series described in Equation 7; 2) upper bound computed using the algorithms presented in Section 4.2; 3) brute-force diagonalization of both the target and gadget Hamiltonian to get the exact eigenvalues. The third option is impractical for general quantum systems of many qubits due to the exponential size of the Hilbert space, though it provides the exact spectral error. The first option is computationally trivial but yields extremely large assignments of Δ (Figure 6). Our algorithm strikes a balance between the two cases by avoiding intense computation while generating Δ assignments that
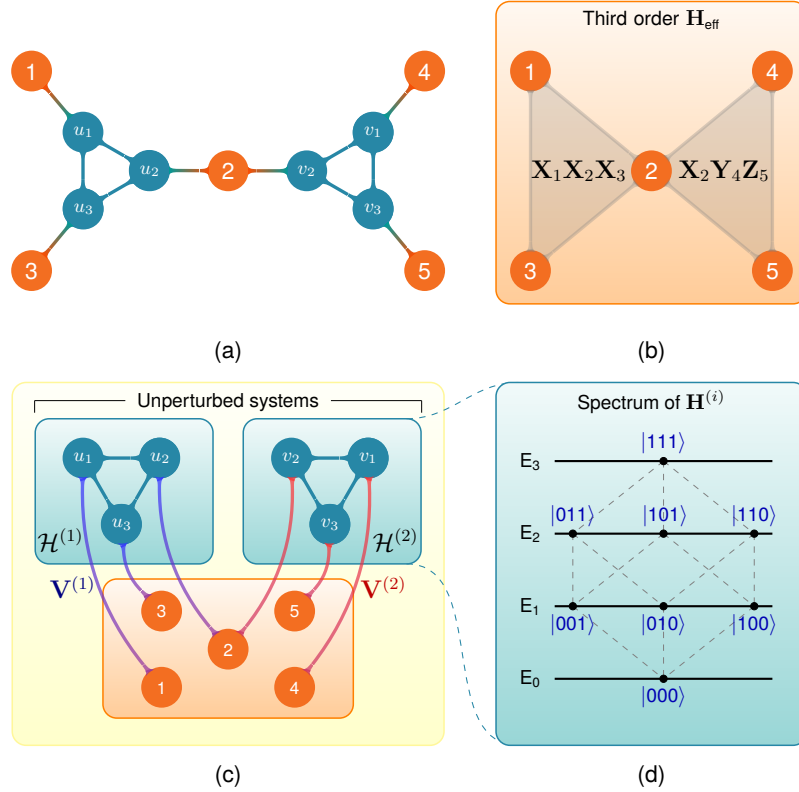
Fig. 4. A numerical example for demonstrating our algorithm estimating the perturbative error. (a) The 11-spin gadget Hamiltonian. Each node corresponds to a spin-1/2 particle and each edge represents an interaction term in the Hamiltonian between two spins. (b) The target 3-body Hamiltonian $H_{\text{eff}} = \alpha_1 X_1 X_2 X_3 + \alpha_2 X_2 Y_4 Z_5$. (c) Rearranging and partitioning the system in (a) according to the setting of perturbation theory used. Here each unperturbed system $H^{(i)}$ consists of three ferromagnetically interacting spins. (d) Spectrum of each subsystem $H^{(i)}$ in (a), $i \in \{1, 2\}$. Here each node represents an eigenstate of $H^{(i)}$. Nodes on a same horizontal dashed line belong to the same energy subspace $\mathcal{P}_j$. There is an edge (ϕ_1, ϕ_2) iff $\|\langle\phi_1|V|\phi_2\rangle\| \neq 0$. For example, if we consider this diagram as representing $H^{(1)}$, since $V^{(1)}|001\rangle_{u_1 u_2 u_3} \propto (|101\rangle + |011\rangle + |000\rangle)_{u_1 u_2 u_3}$ we connect the $|001\rangle$ with the nodes representing $|101\rangle$, $|011\rangle$ and $|000\rangle$.

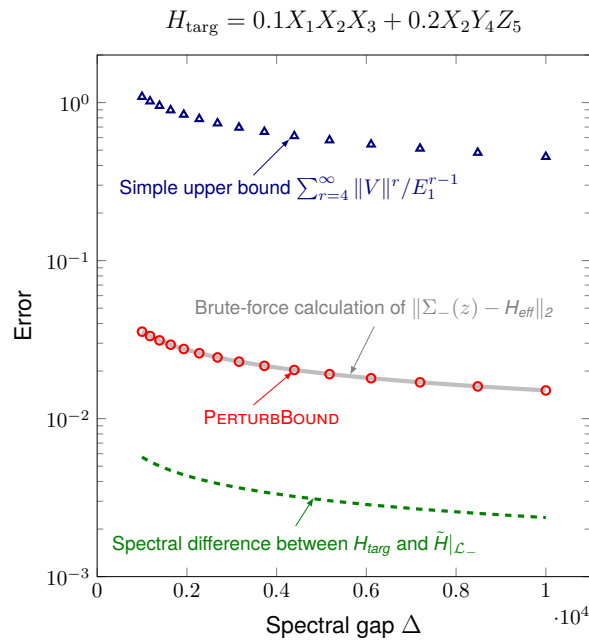$$H_{\text{targ}} = 0.1X_1X_2X_3 + 0.2X_2Y_4Z_5$$

Fig. 5. Comparison between the upper bounds computed using the PERTURBBOUND and the norm computed using (inefficient) explicit matrix-matrix multiplication. The "actual spectral error" in this plot shows the maximum difference between the eigenvalues of H_{eff} and their counterparts in $\tilde{H}$, which are the energies of its 2^N lowest eigenstates with $N = 5$ being the number of particles that H_{eff} acts on (Figure 4b). The actual spectral error is always lower than the error computed based on $\|\Sigma_-(z) - H_{\text{eff}}\|_2$ because $\|\Sigma_-(z) - H_{\text{eff}}\|_2 \leq \epsilon$ is only a *sufficient* condition that guarantees the spectral difference between $\tilde{H}$ and H_{eff} being within ϵ (see [28, Theorem 1]).
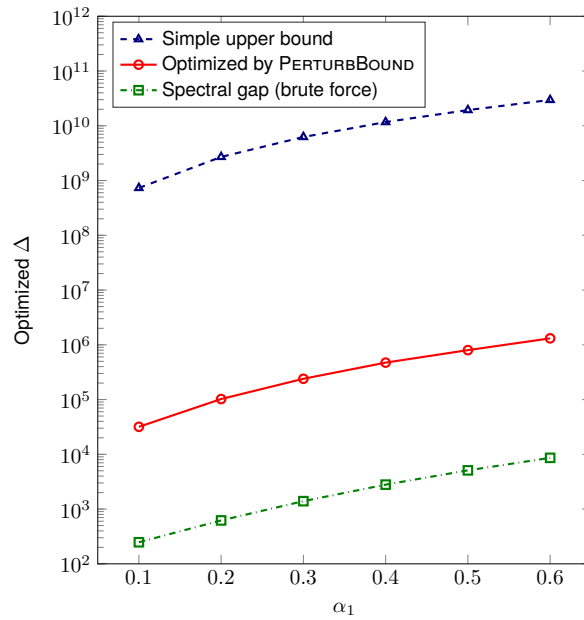
Fig. 6. Comparison between the optimized Δ based on crude error bounds using geometric series ("simple upper bound" in Figure 5), the optimized Δ based on the PERTURBBOUND algorithm presented in Section 4.2 and optimized Δ based on spectral error between the target and the gadget Hamiltonian computed by brute-force diagonalizing both Hamiltonians. Here we use the target Hamiltonian in Figure 4b with $\alpha_2 = 0.2$ fixed and α_1 varying from 0.1 to 0.6. It can be observed that our algorithm significantly improves the assignments for Δ. The gap between the brute force case and the PERTURBBOUND case is due to the fact that $\|\Sigma_-(z) - H_{\text{eff}}\|_2 \leq \epsilon$ is only a *sufficient* condition that guarantees the spectral error to be within ϵ.

are orders of magnitude more practical than the first alternative.

6.2   *Error bounds based on Feynman-Dyson (FD) and Schrieffer-Wolff (SW) series*

In Section 5 we showed that the Schrieffer-Wolff expansion includes more terms than Feynman-Dyson series. Could one potentially improve estimation on the spectral error by adopting Schrieffer-Wolff instead of Feynman-Dyson formalism? Here we show numerical evidence that one indeed could signaficantly improve the error bound. Consider an example where the target Hamiltonian is $H_{\text{targ}} = \alpha X_1 X_2 X_3$ with $\alpha = 0.1$ and the gadget Hamiltonian $\tilde{H} = H + V$ is constructed by adding three ancilla qubits u_1, u_2 and u_3 and defining the Hamiltonians as the following:

$$
\begin{aligned}
H &= \frac{\Delta}{4}(3\mathbf{I} - Z_{u_1} Z_{u_2} - Z_{u_1} Z_{u_3} - Z_{u_2} Z_{u_3}) \\
V &= \mu(X_1 X_{u_1} + X_2 X_{u_2} + X_3 X_{u_3}).
\end{aligned}
\tag{50}
$$

where $\mu = \left(\frac{\alpha\Delta^2}{6}\right)^{1/3}$. Because H_{targ} is 3-body our effective Hamiltonian is truncated at the third order and the remaining terms in the expansion are considered as error:

$$
\begin{aligned}
\Sigma_-(z) = & \underbrace{\frac{3\mu^2}{z-\Delta}\Pi_- + \frac{\Delta^2}{(z-\Delta)^2}\alpha X_1 X_2 X_3 \otimes (|000\rangle\langle 111|_{u_1 u_2 u_3} + |111\rangle\langle 000|_{u_1 u_2 u_3})}_{H_{\text{eff}}=T_1(z)+T_2(z)+T_3(z)} \\
& + \ T_4(z) + T_5(z) + \cdots .
\end{aligned}
\tag{51}
$$

Applying Schrieffer-Wolff transformation to the gadget Hamiltonian yields the low-energy effective Hamiltonian

$$
H_{\text{eff,SW}} = \underbrace{b_1 \Pi_- \hat{R}_1(V_{od})\Pi_-}_{2^{\text{nd}}\text{ order}} + \underbrace{b_1 \Pi_- \hat{R}_2(V_{od})\Pi_-}_{3^{\text{rd}}\text{ order}} + \underbrace{b_1 \Pi_- \hat{R}_3(V_{od})\Pi_- + b_3 \Pi_- \hat{R}_1^3(V_{od})\Pi_-}_{4^{\text{th}}\text{ order}} + \cdots .
\tag{52}
$$

Because the ground state energy of the unperturbed Hamiltonian is 0, the zeroth order term in the expansion (52) vanishes. From Equation 50 the projection of V in the low energy subspace $\mathcal{L}_-$ is 0, thus the first order term also vanishes. The second order term could be rearranged as $-b_1 \Pi_- \hat{V}_{od} R_1 \Pi_-$, which according to Theorem 3 is equivalent to the second order term in the Feynman-Dyson series in Equation 51 for $z \to 0$. At third order, Schrieffer-Wolff expansion gives $b_1 \Pi_- \hat{V}_{od} \mathcal{K} \hat{V}_d R_1 \Pi_-$. Applying Theorem 3 with $r = 3$ we see that this is equivalent to the third order term in the Feynman-Dyson series. Hence up to third order, both formalisms of perturbation theory match up. However, at the fourth order, which is the leading term for the error, difference between the two formalisms starts to show. From the recursive relationship for R_n in Equation 34 we see that R_3 contains a term $-\mathcal{K}\hat{V}_d(R_2)$. So the fourth order term in Equation 52 must contain a term

$$
- b_1 \Pi_- \hat{V}_{od}(-\mathcal{K}\hat{V}_d(R_2))\Pi_- = -b_1 \Pi_- \hat{V}_{od}(-\mathcal{K}\hat{V}_d)^2 R_1 \Pi_-,
\tag{53}
$$

which is equivalent to the entire fourth-order term of the Feynman-Dyson series (Theorem 3 with $r = 4$). The other terms at the fourth order in Equation 52 are beyond Feynman-Dyson series. For example the second term at the fourth order $b_3 \Pi_- \hat{R}_1^3(V_{od})\Pi_-$ corresponds to virtual transitions that switches between $\mathcal{L}_-$ and $\mathcal{L}_+$ multiple times. This violates conditions 1 and 3 in Section 4.1 for sequences $(\phi_0, \cdots, \phi_r)$ that contribute non-trivially to T_r, which results in such terms being excluded from the Feynman-Dyson series.

For varying values of Δ, we calculate the error estimates based on both formulations of perturbation theory and compare them in Figure 7. We have also explicitly diagonalized the target and gadget Hamiltonian and plotted the difference between the low-lying energy levels. The results in Figure 7 shows that Schrieffer-Wolff perturbation theory clearly yields tighter error bounds. The error bounds using Schrieffer-Wolff transformation in Figure 7 are computed by explicit enumeration and evaluation of the terms in the perturbative expansion following Section 5, which is clearly not scalable due to the exponential size of the Hilbert space. The algorithms that we have developed in Section 4.2 could efficiently bound only a subset of the terms in the Schrieffer-Wolff series, namely those of the form in Theorem 3. Bounding the remaining terms in the Schrieffer-Wolff series with similar effectiveness as our algorithms for Feynman-Dyson series requires additional insight and is beyond the scope of our present study.
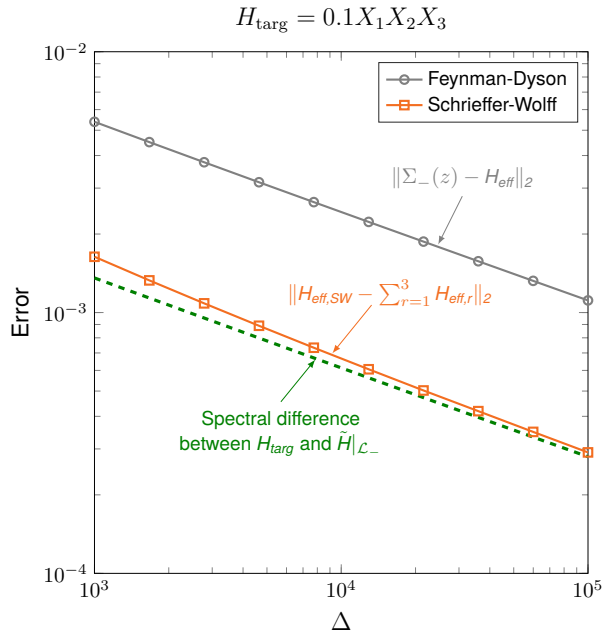
Fig. 7. Comparison between the error bounds computed based on Feynman-Dyson series (Section 4) and Schrieffer-Wolff transformation (Section 5). Here we also show the maximum difference between the lowest $2^3 = 8$ energy levels of the gadget Hamiltonian and the corresponding level of the target Hamiltonian.

7   Summary and conclusion

Perturbative gadgets are the only technique available (as of now and as far as the authors are concerned) for reducing arbitrary many-body Hamiltonian to two-body ones. One of the disadvantages of this technique is the large energy gap Δ needed in the construction of the gadget Hamiltonian, rendering it unnatural in the context of physical systems. Here we address this issue by considering the optimization problem of finding the minimum value of Δ that yields error no greater than a prescribed threshold ϵ (Figure 1). A crucial component of this optimization program is to find tight upper bounds to error terms arising at arbitrary order perturbation theory. In this sense our work is a generalization of [24] to include the gadget constructions in [18, 21].

The problem of computing the error exactly is hard in general because of the exponential size of the Hilbert space. Alternatively, crude upper bounds are trivially attainable via for instance submultiplicativity of operators ($\|AB\| \leq \|A\| \cdot \|B\|$). These bounds are hardly useful for the purpose of *optimizing* the gadget parameters. However, by exploiting the structure of the Hamiltonian we are able to find error bounds that are both orders of magnitude tighter than the crude alternatives (Section 6.1) and efficiently computable (Section 4.2). Each term in the perturbative expansion at a given order is a summation of exponentially many terms. We start from reducing the size of the set of summation from exponential to polynomial in the number of ancilla registers by taking advantage of the structure in the perturbation. We show that there is a hierarchy of equivalence classes (Section 4.1) that allows us to accomplish the reduction. The algorithms for computing the error bounds presented in Section 4.2 take

advantage of such hierarchical structure. In the special case where the target terms $H_{\mathrm{targ},i}$ pairwise commute, we show that our error bounds are sharp (Section 4.3).

However, a gap still exists (Figure 6) between the output of our algorithm and the result of brute-force optimization. This gap is due to the machinery of perturbation theory that we use (illustrated in Figure 5), which is based on the Feynman-Dyson series. In Section 6.2 we observe numerically that using the Schrieffer-Wolff transformation [32, 31] instead may enable one to get closer to the brute-force results (Figure 7). This improvement may be explained by Theorem 3 which says that a specific class of terms in the Schrieffer-Wolff series already captures all of the terms in the Feynman-Dyson series (Section 5). It is tempting to consider whether our technique can be applied to obtain efficient error bounds for Schrieffer-Wolff series. One challenge in this regard is that our efficient algorithm is built on the observation that the terms at each order is essentially a summation of walks in the eigenspace of the unperturbed Hamiltonian, per Equation 9. This combinatorial picture of summing over walks comes from the matrix product structure of the self-energy expansion (Equation 1). Whether this same structure exists in Schrieffer-Wolff transformation (and other formalisms of perturbation theory) remains to be assessed.

Acknowledgments

References

1. A. Yu. Kitaev. Fault-tolerant quantum computation by anyons. *Annals of Physics*, 303(1):2 – 30, 2003. arXiv:quant-ph/9707021.
2. Chetan Nayak, Steven H. Simon, Ady Stern, Michael Freedman, and Sankar Das Sarma. Non-abelian anyons and topological quantum computation. *Rev. Mod. Phys.*, 80:1083–1159, Sep 2008. arXiv:0707.1889 [cond-mat.str-el].
3. Michael H. Freedman, Alexei Kitaev, Michael J. Larsen, and Zhenghan Wang. Topological quantum computation, 2001. arXiv:quant-ph/0101025.
4. R. Walter Ogburn and John Preskill. Topological quantum computation. In *Selected Papers from the First NASA International Conference on Quantum Computing and Quantum Communications*, QCQC '98, pages 341–356, London, UK, UK, 1998. Springer-Verlag. 10.1.1.37.4281.
5. Robert Raussendorf and Hans J. Briegel. A one-way quantum computer. *Phys. Rev. Lett.*, 86:5188–5191, May 2001. 10.1103/PhysRevLett.86.5188.
6. Robert Raussendorf, Daniel E. Browne, and Hans J. Briegel. Measurement-based quantum computation on cluster states. *Phys. Rev. A*, 68:022312, Aug 2003. 10.1103/PhysRevA.68.022312.
7. Hans J. Briegel and Robert Raussendorf. Persistent entanglement in arrays of interacting particles. *Phys. Rev. Lett.*, 86:910–913, Jan 2001. arXiv:quant-ph/0004051.
8. Robert Raussendorf, Sergey Bravyi, and Jim Harrington. Long-range quantum entanglement in noisy cluster states. *Phys. Rev. A*, 71:062313, Jun 2005. arXiv:quant-ph/0407255.
9. Dan E. Browne and Hans J. Briegel. One-way quantum computation - a tutorial introduction, 2006. arXiv:quant-ph/0603226.
10. Robert Raussendorf, Daniel Browne, and Hans Briegel. The one-way quantum computer–a non-network model of quantum computation. *Journal of Modern Optics*, 49(8):1299–1306, 2002. arXiv:quant-ph/0603226.
11. Ryan Babbush, Peter J. Love, and Alán Aspuru-Guzik. Adiabatic quantum simulation of quantum

chemistry. *Scientific Reports*, 4(6603), 2014. arXiv:1311.3967 [quant-ph].

12. Ari Mizel, Daniel A. Lidar, and Morgan Mitchell. Simple proof of equivalence between adiabatic quantum computation and the circuit model. *Phys. Rev. Lett.*, 99:070502, Aug 2007. 10.1103/PhysRevLett.99.070502.

13. Seth Lloyd and Barbara M Terhal. Adiabatic and hamiltonian computing on a 2d lattice with simple two-qubit interactions. *New Journal of Physics*, 18(2):023042, 2016. arXiv:1509.01278 [quant-ph].

14. David Gosset, Barbara M. Terhal, and Anna Vershynina. Universal adiabatic quantum computation via the space-time circuit-to-hamiltonian construction. *Phys. Rev. Lett.*, 114:140501, Apr 2015. 10.1103/PhysRevLett.114.140501.

15. Dorit Aharonov, Wim van Dam, Julia Kempe, Zeph Landau, Seth Lloyd, and Oded Regev. Adiabatic quantum computation is equivalent to standard quantum computation. *Society for Industrial and Applied Mathematics Journal of Computing*, 37:166–194, 2007. arXiv:quant-ph/0405098.

16. M. Stewart Siu. Adiabatic rotation, quantum search, and preparation of superposition states. *Phys. Rev. A*, 75:062337, Jun 2007. 10.1103/PhysRevA.75.062337.

17. A. Kitaev, A. H. Shen, and M. N. Vyalyi. *Classical and Quantum Computation*. AMS Graduate Studies in Mathematics, 2002.

18. J. Kempe, A. Kitaev, and O. Regev. The Complexity of the Local Hamiltonian Problem. *SIAM J. Computing*, 35(5):1070–1097, 2006. arXiv:quant-ph/0406180.

19. R. Oliveira and B. Terhal. The complexity of quantum spin systems on a two-dimensional square lattice. *Quant. Inf. and Comp.*, 8(10):0900–0924, 2008. arXiv:quant-ph/0504050.

20. S. Bravyi, D. DiVincenzo, D. Loss, and B. Terhal. Quantum Simulation of Many-Body Hamiltonians Using Perturbation Theory with Bounded-Strength Interactions. *Phys. Rev. Lett.*, 101:070503, 2008. arXiv:0803.2686v1.

21. S. P. Jordan and E. Farhi. Perturbative gadgets at arbitrary orders. *Phys. Rev. A*, 062329, 2008. arXiv:0802.1874v4.

22. S. Bravyi, D. DiVincenzo, R. Oliveira, and B. Terhal. The Complexity of Stoquastic Local Hamiltonian Problems. *Quant. Inf. and Comp.*, 8(5), 2006. quant-ph/0606140.

23. Toby S. Cubitt and Ashley Montanaro. Complexity classification of local hamiltonian problems. In *Proceedings of the 2014 IEEE 55th Annual Symposium on Foundations of Computer Science*, FOCS '14, pages 120–129, Washington, DC, USA, 2014. IEEE Computer Society.

24. Yudong Cao, Ryan Babbush, Jacob Biamonte, and Sabre Kais. Hamiltonian gadgets with reduced resource requirements. *Phys. Rev. A*, 91(1):012315, 2015. arXiv:1311.2555 [quant-ph].

25. J. D. Biamonte. Non-perturbative k-body to two-body commuting conversion Hamiltonians and embedding problem instances into Ising spins. *Phys. Rev. A*, 77(5):052331, 2008. arXiv:0801.3800.

26. Samuel A. Ocko and Beni Yoshida. Nonperturbative Gadget for Topological Quantum Codes. *Phys. Rev. Lett.*, 107(250502), 2011. arXiv:1107.2697 [quant-ph].

27. Yiğit Subaşı and Christopher Jarzynski. Nonperturbative embedding for highly nonlocal Hamiltonians. *Phys. Rev. A*, 94:012342, Jul 2016. arXiv:1601.02922 [quant-ph].

28. Yudong Cao and Sabre Kais. Efficient estimation of perturbative error with cellular automata. 2016. arXiv:1607.01374 [quant-ph].

29. Yudong Cao. *Combinatorial algorithms for perturbation theory and applications on quantum computing*. PhD thesis, Purdue University, West Lafayette, IN, August 2016. Available online.

30. Alexander L. Fetter and John Dirk Walecka. *Quantum Theory of Many-Particle Systems*. McGraw-Hill, New York, 1971.

31. S. Bravyi, D. DiVincenzo, and D. Loss. Schrieffer-Wolff transformation for quantum many-body systems. *Ann. Phys.*, 326(10), 2011. arXiv:1105.0675.

32. J. R. Schrieffer and P. A. Wolff. Relation between the Anderson and Kondo Hamiltonians. *Phys. Rev.*, 149:491–492, Sep 1966. 10.1103/PhysRev.149.491.

33. Claude Bloch. Sur la théorie des perturbations des états liés. *Nuclear Physics*, 6:329–347, 1958.

34. R. Shankar. *Principles of Quantum Mechanics*. Springer US, 1994.

35. Yudong Cao and Daniel Nagaj. Perturbative gadget without strong interactions. *Quantum Infor-*

mation and Computation, 15:1197–1222, 2015. arXiv:1408.5881 [quant-ph].

36. J. Goldstone. Derivation of the brueckner many-body theory. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 239(1217):267–279, 1957.

37. Let $f_{m\ell}$ be the total number of possible reduced configurations of length m and maximum possible number of energy levels ℓ. Then $f_{m\ell} \leq m^{\ell}$ for any $m \geq 2$ and $\ell \geq 1$. Proof: The last element of a reduced configuration could take any one of ℓ values. Without loss of generality, assume the elements of a reduced configuration is non-decreasing, the remaining $m-1$ elements of $\tilde{\mathbf{c}}$ has $f_{m-1,\tilde{\mathbf{c}}_m}$ choices where $\tilde{\mathbf{c}}_m \in \{0, \cdots, \ell-1\}$ is the last element of $\tilde{\mathbf{c}}$. We then have the recursion $f_{m\ell} = f_{m-1,\ell} + f_{m-1,\ell-1} + \cdots + f_{m-1,1}$ with boundary condition $f_{k1} = 1$ for any $k \in \{1, \cdots, m\}$ and $f_{1k} = k$ for any $k \in \{0, 1, \cdots, \ell-1\}$. Hence $f_{m\ell} = f_{m-1,\ell} + f_{m,\ell-1} = 1 + \sum_{i=1}^{m} f_{i,\ell-1}$. Starting from $f_{m1} = 1$, we have $f_{m2} = 1 + f_{11} + f_{21} + \cdots + f_{m1} \leq 1 + mf_{m1} = 1 + m$ and $f_{m3} = 1 + f_{12} + f_{22} + \cdots + f_{m2} \leq 1 + m + m^2$. Applying this to $f_{m\ell}$, we have $f_{m\ell} \leq 1 + f_{m,\ell-1} \leq 1 + m(1 + mf_{m,\ell-2}) \leq \cdots \leq 1 + m + \cdots + m^{\ell-1} \leq m^{\ell}$.

Appendix A: An example for illustrating notions introduced in Section 4.1

This example is essentially the one considered in Section 6 but here we abstract out only the revelant aspects of the example without going into full detail. Suppose our target Hamiltonian H_{targ} is a sum of two 3-local terms that need to be reduced to 2-local using the gadget construction (Section 2). Our gadget Hamiltonian $\tilde{H} = H + V$ has the unperturbed part $H = H^{(1)} + H^{(2)}$ acting on two registers of three ancilla qubits (because the target terms are 3-local). The perturbation V couples to each register of ancillas with interaction strengths λ_1 and λ_2 (as a reminder, see Equation 5 and the restriction that $\lambda_{i,j} = \lambda_i$ introduced at the beginning of Section 4.1). Hence $m = 2$ and $k = 3$ in this example and the low energy level of H satisfies $E(\phi) = E^{(1)}(j_1) + E^{(2)}(j_2) = 0$ with j_1, j_2 being either 0 or 3. At second order, from previous discussion we see that the sequences of reduced configurations that contribute non-trivially to T_2 are

$$\tilde{\mathbf{c}}_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \rightarrow \tilde{\mathbf{c}}_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \rightarrow \tilde{\mathbf{c}}_2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\tilde{\mathbf{c}}_0 = \begin{pmatrix} 0 \\ 3 \end{pmatrix} \rightarrow \tilde{\mathbf{c}}_1 = \begin{pmatrix} 1 \\ 3 \end{pmatrix} \rightarrow \tilde{\mathbf{c}}_2 = \begin{pmatrix} 0 \\ 3 \end{pmatrix}$$

$$\tilde{\mathbf{c}}_0 = \begin{pmatrix} 0 \\ 3 \end{pmatrix} \rightarrow \tilde{\mathbf{c}}_1 = \begin{pmatrix} 0 \\ 2 \end{pmatrix} \rightarrow \tilde{\mathbf{c}}_2 = \begin{pmatrix} 0 \\ 3 \end{pmatrix} \tag{A.1}$$

$$\tilde{\mathbf{c}}_0 = \begin{pmatrix} 3 \\ 3 \end{pmatrix} \rightarrow \tilde{\mathbf{c}}_1 = \begin{pmatrix} 2 \\ 3 \end{pmatrix} \rightarrow \tilde{\mathbf{c}}_2 = \begin{pmatrix} 3 \\ 3 \end{pmatrix}.$$

Accordingly, the set $\mathcal{W}_2^{\mathbf{c}}$ consists of the following sequences of configurations

$$\mathbf{c}_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \rightarrow \mathbf{c}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \rightarrow \mathbf{c}_2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}; \quad \mathbf{c}_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \rightarrow \mathbf{c}_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \rightarrow \mathbf{c}_2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix};$$

$$\mathbf{c}_0 = \begin{pmatrix} 0 \\ 3 \end{pmatrix} \rightarrow \mathbf{c}_1 = \begin{pmatrix} 1 \\ 3 \end{pmatrix} \rightarrow \mathbf{c}_2 = \begin{pmatrix} 0 \\ 3 \end{pmatrix}; \quad \mathbf{c}_0 = \begin{pmatrix} 3 \\ 0 \end{pmatrix} \rightarrow \mathbf{c}_1 = \begin{pmatrix} 3 \\ 1 \end{pmatrix} \rightarrow \mathbf{c}_2 = \begin{pmatrix} 3 \\ 0 \end{pmatrix};$$

$$\mathbf{c}_0 = \begin{pmatrix} 0 \\ 3 \end{pmatrix} \rightarrow \mathbf{c}_1 = \begin{pmatrix} 0 \\ 2 \end{pmatrix} \rightarrow \mathbf{c}_2 = \begin{pmatrix} 0 \\ 3 \end{pmatrix}; \quad \mathbf{c}_0 = \begin{pmatrix} 3 \\ 0 \end{pmatrix} \rightarrow \mathbf{c}_1 = \begin{pmatrix} 2 \\ 0 \end{pmatrix} \rightarrow \mathbf{c}_2 = \begin{pmatrix} 3 \\ 0 \end{pmatrix}; \tag{A.2}$$

$$\mathbf{c}_0 = \begin{pmatrix} 3 \\ 3 \end{pmatrix} \rightarrow \mathbf{c}_1 = \begin{pmatrix} 2 \\ 3 \end{pmatrix} \rightarrow \mathbf{c}_2 = \begin{pmatrix} 3 \\ 3 \end{pmatrix}; \quad \mathbf{c}_0 = \begin{pmatrix} 3 \\ 3 \end{pmatrix} \rightarrow \mathbf{c}_1 = \begin{pmatrix} 3 \\ 2 \end{pmatrix} \rightarrow \mathbf{c}_2 = \begin{pmatrix} 3 \\ 3 \end{pmatrix}.$$

Note that the configuration sequences on each row of (A.2) is formed by permuting elements of the reduced configuration sequence on the corresponding row in (A.1). Finally, each configuration sequence in (A.2) can be replaced with sequences of states, forming the set $\mathcal{W}_2$ which consists of the following sequences of states $\phi_0 \to \phi_1 \to \phi_2$ (Here | separates the two ancilla registers and each block of 3 sequences corresponds to the configuration sequence in the associated row and position in (A.2)):

$$
\begin{array}{ll}
000|000 \to 100|000 \to 000|000 & 000|000 \to 000|100 \to 000|000 \\
000|000 \to 010|000 \to 000|000 & 000|000 \to 000|010 \to 000|000 \\
000|000 \to 001|000 \to 000|000 & 000|000 \to 000|001 \to 000|000 \\[6pt]
000|111 \to 100|111 \to 000|111 & 111|000 \to 111|100 \to 111|000 \\
000|111 \to 010|111 \to 000|111 & 111|000 \to 111|010 \to 111|000 \\
000|111 \to 001|111 \to 000|111 & 111|000 \to 111|001 \to 111|000 \\[6pt]
000|111 \to 000|011 \to 000|111 & 111|000 \to 011|000 \to 111|000 \\
000|111 \to 000|101 \to 000|111 & 111|000 \to 101|000 \to 111|000 \\
000|111 \to 000|110 \to 000|111 & 111|000 \to 110|000 \to 111|000 \\[6pt]
111|111 \to 011|111 \to 111|111 & 111|111 \to 111|011 \to 111|111 \\
111|111 \to 101|111 \to 111|111 & 111|111 \to 111|101 \to 111|111 \\
111|111 \to 110|111 \to 111|111 & 111|111 \to 111|110 \to 111|111
\end{array}
\tag{A.3}
$$

We observe that the first (top left) block of three sequences in (A.3) sums up to a term $3\lambda_1^2 \cdot \frac{1}{z-E_1}|000\rangle\langle 000|$, with each sequence contributing a term $\lambda_1 \cdot \frac{1}{|z-E_1|} \cdot \lambda_1$ in the final upper bound (Equation 18). Similarly we see that the top right block of (A.3) sums up to $3\lambda_2^2 \cdot \frac{1}{z-E_1}|000\rangle\langle 000|$. Recall that Π_- is the projector onto the low energy subspace $\mathcal{L}_- = \mathcal{L}_-^{(1)} \otimes \mathcal{L}_-^{(2)}$ with each $\mathcal{L}_-^{(i)} = \text{span}\{|000\rangle, |111\rangle\}$. Adding up the terms in all the sequences in (A.3) gives a term $3(\lambda_1^2 + \lambda_2^2) \cdot \frac{1}{z-E_1}\Pi_-$, which is symmetric with respect to the permutation of registers.

Also observe that we are able to calculate the coefficient $3(\lambda_1^2 + \lambda_2^2) \cdot \frac{1}{z-E_1}$ in the sum over *all* blocks of sequences in (A.3) by only inspecting the first row, gleaning two terms with coefficients $3\lambda_1^2 \cdot \frac{1}{z-E_1}$ and $3\lambda_2^2 \cdot \frac{1}{z-E_1}$. This is because the set $\mathcal{W}_r$ is invariant with respect to the operation of flipping all the bits of any set of registers (recall discussion prior to Equation 15). Examining A.3 one could find that for instance flipping all the bits in the first register of the top left block yields the block on the right of third row. Flipping all the bits in the second register of the top left block yields the block on the left of the second row. Flipping all the bits in both registers yields the bottom right block, etc. Therefore in order to find the coefficients to T_2 it suffices to focus on only the sequences $\phi_0 \to \phi_1 \to \phi_2$ where $\phi_0 = 000|000$ (Equation 15).