Zero-Overhead Cancellation of the Leading Trotter Error via Commutator-Graph Scheduling

Molena Huynh*
North Carolina State University, Raleigh, North Carolina 27695, USA
(Dated: July 1, 2026)

A quantum computer is a stochastic device: the Born rule returns one bitstring per run, so every quantity a simulation reports—fidelities, observables, and the very error one is trying to control—is a statistical estimate over samples, and the structure that governs those estimates is the algebra of the operators being evolved. We introduce *commutator-graph scheduling* (CGS), which doubles the convergence order of first-order Trotterization at zero gate, qubit, or depth overhead by making the leading error an explicit, estimable operator on that algebra. The bottleneck is well known: term order within a Lie–Trotter step is free, yet ordering heuristics buy only marginal accuracy and no one had explained why. CGS recasts the leading error as an exact, matrix-free operator, the *commutator error form* $E_\pi = \sum_{a<b}[h_{\pi(a)}, h_{\pi(b)}]$—a signed sum of Pauli strings over the edges of the commutator graph, validated against dense exponentials to 10^{-9}. We prove that reordering only re-signs E_π, leaving its norm exactly invariant on collision-free families, so no fixed ordering escapes the $O(t^2/r)$ rate. The true free parameter is the per-step *schedule*: a step-alternating *antithetic* schedule cancels E_π and converges at $O(t^3/r^2)$ with the identical $L \cdot r$ rotations. Across 144 structured Hamiltonians versus exact statevector references, CGS lifts the empirical rate from 1.96 to 4.07, reaches fidelity 0.99 with 1.40× fewer rotations than the best first-order ordering, and attains 0.999 where first-order Trotter cannot. A matrix-free Krylov reference extends the exact benchmark to 14 qubits, where the order doubling is size independent ($\approx$2 vs. $\approx$4 from 4 to 14 qubits) and the matched-cost speedup reaches 2.06×. An interpretable scheduler over ten graph features selects the gate-optimal schedule per instance with 93% leave-one-out accuracy. CGS is a drop-in, fully reproducible upgrade for near-term quantum simulation.

I. INTRODUCTION

Statistics is not an afterthought in quantum computing; it is structural to it. Seen from the keyboard, a quantum computer is a stochastic device interrogated only by sampling: the Born rule means each run returns one bitstring drawn from a distribution, so everything a simulation ultimately reports is an estimate—observables carry variance and confidence intervals, and the residual error of an approximate circuit is itself something one can only *measure* to finite precision. This viewpoint sharpens the problem of simulating time evolution. The error one is trying to suppress is a concrete operator on the Pauli algebra, and controlling it well means representing it exactly, bounding it rigorously, and estimating its magnitude from shots with a stated risk—a chain of statistical reasoning that runs from the operator algebra to the sampled witness. Commutator-graph scheduling, introduced below, follows exactly this chain: it turns the leading Trotter error into an explicit, matrix-free operator on a graph, proves how far one can and cannot shrink it, and gives an unbiased shot-based estimator of its size so the accuracy claim is empirically certifiable rather than merely asymptotic.

The simulation of quantum time evolution $e^{-\mathrm{i}Ht}$ is central to a wide array of fields, such as quantum chemistry, condensed-matter physics, and materials science, and is among the oldest and most promising applications of quantum computers [1, 2, 22, 23]. Its appeal extends from quantum chemistry, where simulated time evolution underlies phase estimation of molecular energies [23, 24], to the first digital quantum simulations on trapped-ion and other near-term hardware [25, 26]. The workhorse primitive for this task is the product formula. Writing a Hamiltonian in its Pauli decomposition,

$$H = \sum_{j=1}^{L} c_j P_j, \tag{1}$$

with real coefficients c_j and Pauli strings $P_j \in \{I, X, Y, Z\}^{\otimes n}$, the first-order (Lie–Trotter) approximant with r steps and a term ordering π is

$$U_{\mathrm{Trot}}(\pi) = \left(\prod_{j \in \pi} e^{-\mathrm{i}\, c_j P_j\, t/r} \right)^r. \tag{2}$$

The accuracy of this approximation is governed by the commutators of the summands: the leading correction to a single step is $\frac{t^2}{2r} \sum_{a<b}[c_a P_a, c_b P_b]$, which vanishes for any pair of terms that commute [3–5]. The product-formula construction itself traces to the operator-splitting schemes of Strang [27] and to Suzuki's fractal decompositions [4, 28], and it remains competitive with the more recent quantum-signal-processing and linear-combination-of-unitaries primitives [30, 32, 34, 35] precisely because its cost scales with commutators rather than with a global norm. The modern, tight analysis of this error in terms of nested commutators provides the theoretical anchor of the present work [5, 6]. This

* molena.huynh@jmp.com

commutator-scaling viewpoint also explains why product formulas are nearly optimal for geometrically local lattice Hamiltonians [31, 36].

A defining feature of the first-order formula is that its implementation cost is one rotation $e^{-\mathrm{i}c_j P_j t/r}$ per term per step, so the gate count is $L \cdot r$ *independently* of the ordering π, while the error is not. The ordering is therefore a free knob, and a substantial body of empirical work has asked how much accuracy it can buy [10, 11, 15, 48]. A parallel line of work has shown that the worst-case Trotter bound is often loose: errors from successive steps can interfere destructively [37, 38], are bounded by quantum localization for local observables [41, 42], and improve markedly on average over input states [39, 40]. The natural structure for this question is the *commutator graph* (or anticommutation graph), with one node per term and an edge between every pair of terms that anticommute—the same graph used in measurement grouping for variational algorithms, where a clique cover into mutually commuting sets reduces the number of measurement settings [12–14, 55, 56]. Despite this rich structure, the empirical verdict has been consistently underwhelming: ordering heuristics move the error only marginally, and their gains are widely regarded as "within noise." *Why* a free parameter should be so impotent has not been explained.

We resolve this, and turn it into an asset, by introducing *commutator-graph scheduling* (CGS). At the center of CGS is the *commutator error form*, the exact leading-order error operator

$$E_\pi = \sum_{a<b} \big[h_{\pi(a)}, h_{\pi(b)}\big], \qquad h_j = c_j P_j, \tag{3}$$

which we assemble without ever forming a $2^n \times 2^n$ matrix, as a signed sum of Pauli strings indexed by the edges of the commutator graph (Sec. II). Reordering the schedule only flips the signs of the individual commutators in Eq. (3); it never changes which commutators appear. The Hilbert–Schmidt norm of E_π is therefore *exactly* ordering invariant whenever the edge commutators land on distinct Pauli strings, and only a small "colliding" fraction is ordering reducible at all (Theorem 1). This is the precise, provable content of the "ordering is within noise" folklore: no fixed ordering escapes the $O(t^2/r)$ first-order rate.

The resolution is to enlarge the free degree of freedom. The order applied at each Trotter step need not be the same; the genuinely free object is a per-step *schedule* $\sigma = (\pi_1, \ldots, \pi_r)$, realized at the identical $L \cdot r$ rotation count. Among all schedules, one cancels Eq. (3) *exactly*: the *antithetic* schedule, which alternates a term order with its reverse on successive steps. Each adjacent step-pair is then a palindromic, and hence symmetric, product formula whose leading $O(t^2/r)$ commutator error vanishes, upgrading the convergence from $O(t^2/r)$ to $O(t^3/r^2)$ at *no change in gate count* (Theorem 2). The underlying mechanism is the classical symmetrization of Suzuki [4, 7]; reading it as a zero-overhead *scheduling* choice—one that the impotence theorem shows a single fixed ordering can never realize—is new, and it changes the practical recommendation for first-order simulation.

This manuscript makes six contributions. We introduce the commutator error form [Eq. (3)], a matrix-free representation of the leading Trotter error as a signed Pauli-string sum over the commutator graph, validated against the dense pair-commutator sum to 10^{-9} (Definition 1). We prove that a single ordering cannot reduce the leading-error norm below an instance floor that is exactly ordering invariant on collision-free families (Theorem 1), which explains the empirical ceiling that ordering heuristics encounter. We prove that the antithetic schedule cancels the leading error and converges at second order at the identical gate budget (Theorem 2), and we combine it with a commuting-clique compression of the commutator graph (Proposition 2). We benchmark CGS against exact statevector references on 144 structured Hamiltonian instances, reporting rotations to a target fidelity and the empirical convergence rate; CGS reaches 0.99 with 1.40× fewer rotations and is the only family of schedules to reach 0.999 within the swept budget. A matrix-free Krylov reference extends the exact benchmark from the dense-exponential ceiling to 14 qubits; across $n = 4$ to 14 the convergence-order doubling (first-order rate ≈ 2, antithetic rate ≈ 4) and the matched-cost speedup are size independent (2.06× at $n = 14$; Sec. IV D). Finally, we add a learned scheduler over ten commutator-graph features that selects the gate-optimal schedule per instance with 93% leave-one-out accuracy and recovers near-oracle efficiency (114.0 rotations vs. the oracle's 112.4) (Proposition 3).

The boundaries of the study are explicit. CGS is evaluated by classical statevector simulation on systems up to $n = 14$ qubits, where an exact reference is tractable, on synthetic and structured Hamiltonians, under a rotation-count cost model; the gains are asymptotic in the step size, and we report the crossover budget. The antithetic schedule is not a new product-formula primitive—the symmetrization is classical—but its certified, zero-overhead deployment on the commutator graph, together with the impotence theorem that motivates it and the learned scheduler that selects it, is new.

This work continues a program on the spectral truncation of operators for query-efficient quantum optimization and simulation. Earlier entries develop graph-conditioned trust regions for budgeted parameter search [57, 58], the measurement and query accounting that fixes the cost model [59, 60], and topology-conditioned transfer of optimizer state across instances [61], culminating in operator-spectral truncated priors and spectral-truncation graph kernels for QAOA warm starts [62, 63]. The present manuscript carries the same operator-spectral viewpoint from variational optimization to Hamiltonian simulation: the commutator error form is a matrix-free, edge-resolved truncation of the leading product-formula error on the commutator graph, and the matched-cost analysis reuses the query-budget

discipline established in that line of work.

II. COMMUTATOR-GRAPH SCHEDULING

We now present the CGS framework and its analytic guarantees. We first record the Pauli algebra that makes an exact, matrix-free simulator and error form possible (Sec. II A), then define the commutator error form (Sec. II B), prove that ordering is impotent (Sec. II C), prove that the antithetic schedule cancels the leading error at the same gate budget (Sec. II D), establish the commuting-clique compression (Sec. II E), and describe the learned scheduler that selects among schedules (Sec. II F). Throughout, every result is stated under explicitly named hypotheses with a complete proof, and every numerical claim is anchored to a quantity that the accompanying code regenerates.

Figure 1 summarizes the pipeline. Each Hamiltonian instance is decomposed into Pauli terms; the commutator graph $\mathcal{G}(H)$ places one node per term and an edge between every anticommuting pair, and a greedy coloring partitions the terms into mutually commuting cliques. A schedule fixes, for every Trotter step, a term order and a fold—first-order, or the step-alternating antithetic fold—while holding the realized rotation count fixed. The product formula is executed by an exact, matrix-free statevector simulator and scored both by fidelity against $e^{-\mathrm{i}Ht}$ and by the rotations needed to reach a target fidelity. We stress that no number is reported until three integrity gates pass: the fast Pauli-rotation kernel and dense `expm` agree to 10^{-9}, the matrix-free error form agrees with the dense pair-commutator sum to 10^{-9}, and the matrix-free Krylov reference (which extends the exact benchmark to 14 qubits) agrees with dense `expm` to 10^{-9}.

A. Pauli algebra and the matrix-free simulator

A Pauli string $P \in \{I, X, Y, Z\}^{\otimes n}$ acts on $|\psi\rangle \in \mathbb{C}^{2^n}$; we write $\{A, B\} = AB + BA$ and $[A, B] = AB - BA$, and use $X^2 = Y^2 = Z^2 = I$ together with $\{X, Y\} = \{Y, Z\} = \{Z, X\} = 0$. Two structural facts make an exact simulator possible without ever forming a $2^n \times 2^n$ matrix.

Lemma 1 (Commutation is an $O(n)$ parity test)**.** *Let $P = \bigotimes_{q=1}^{n} p_q$ and $Q = \bigotimes_{q=1}^{n} q_q$ be Pauli strings. Call position q* anticommuting *if $p_q \neq q_q$ and neither is I, and let $k = k(P, Q)$ count the anticommuting positions. Then $PQ = (-1)^k QP$, so P and Q commute iff k is even. The test runs in $O(n)$ time and forms no $2^n \times 2^n$ matrix.*

Proof. At each position either $p_q = q_q$ or one is I (then $p_q q_q = q_q p_q$), or $p_q \neq q_q$ and neither is I (then $p_q q_q = -q_q p_q$). So $p_q q_q = (-1)^{\epsilon_q} q_q p_q$ with $\epsilon_q = 1$ exactly on anticommuting positions, and since the tensor product factorizes across qubits, $PQ = \big(\prod_q (-1)^{\epsilon_q}\big) QP = (-1)^k QP$ with $k = \sum_q \epsilon_q$. Hence $PQ = QP$ iff k is even. □

Lemma 2 (Pauli rotation as a cosine–sine pair)**.** *For a non-identity Pauli string P ($P^2 = I$) and $\theta \in \mathbb{R}$,*

$$e^{-\mathrm{i}\theta P} = \cos\theta\, I - \mathrm{i}\sin\theta\, P, \tag{4}$$

so $e^{-\mathrm{i}\theta P}|\psi\rangle = \cos\theta\,|\psi\rangle - \mathrm{i}\sin\theta\, P|\psi\rangle$ is one sparse application of P plus a scaled vector add, computable in $O(n\,2^n)$ time without forming the $2^n \times 2^n$ matrix.

Proof. Since $P^{2m} = I$ and $P^{2m+1} = P$, regrouping the convergent series of $e^{-\mathrm{i}\theta P}$ by parity gives $\cos\theta\, I - \mathrm{i}\sin\theta\, P$. A single Pauli string maps each basis state to one basis state with a unit-modulus phase, so $P|\psi\rangle$ costs $O(n\,2^n)$. □

We emphasize that these identities are not assumed but verified: before any fidelity is computed, Eq. (4) is cross-checked against dense `scipy.linalg.expm` on random small Paulis to tolerance 10^{-9} (the `pauli_algebra_verified` flag). The product of two Pauli strings is itself a phased Pauli, $PQ = \omega R$ with $\omega \in \{\pm 1, \pm\mathrm{i}\}$ and R a Pauli string, computed in $O(n)$ (the `pauli_product` kernel); this is the algebra on which the commutator error form rests.

B. The commutator error form

The commutator graph $\mathcal{G}(H)$ has one node per term and an edge $\{j, k\}$ for every anticommuting pair (Lemma 1). The exact value of the pairwise commutators that populate the leading error is fixed by the following lemma.

Lemma 3 (Pauli commutator value)**.** *Let $h_j = c_j P_j$ and $h_k = c_k P_k$ with real c_j, c_k. If P_j, P_k commute then $[h_j, h_k] = 0$; if they anticommute then $[h_j, h_k] = 2c_j c_k\, P_j P_k = 2c_j c_k\, \omega_{jk} R_{jk}$ with $\omega_{jk} \in \{\pm\mathrm{i}\}$ and R_{jk} a Hermitian Pauli string, so $\|[h_j, h_k]\|_{\mathrm{HS}} = 2\,|c_j c_k|\, 2^{n/2}$ and the operator norm is $2\,|c_j c_k|$.*

Proof. By Lemma 1, $P_k P_j = (-1)^{k(P_j, P_k)} P_j P_k$, so the commutator vanishes when the terms commute and equals $2c_j c_k\, P_j P_k$ when they anticommute. The product $P_j P_k = \omega_{jk} R_{jk}$ is a phased Pauli (Sec. II A); because the commutator of Hermitian operators is anti-Hermitian and R_{jk} is Hermitian, $\omega_{jk} \in \{\pm\mathrm{i}\}$. Every Pauli string is unitary with $\|R_{jk}\| = 1$ and $\|R_{jk}\|_{\mathrm{HS}}^2 = \mathrm{Tr}(R_{jk}^2) = 2^n$, which gives the stated norms. □

Definition 1 (Commutator error form)**.** *For an ordering π, the leading error operator is $E_\pi = \sum_{a<b} [h_{\pi(a)}, h_{\pi(b)}]$. By Lemma 3 it equals the edge sum*

$$E_\pi = \sum_{\{j,k\} \in E(\mathcal{G})} \varepsilon_{jk}(\pi)\, 2c_j c_k\, (P_j P_k), \tag{5}$$

a signed sum of phased Pauli strings indexed by the edges of $\mathcal{G}(H)$, with $\varepsilon_{jk}(\pi) = +1$ if j precedes k in π and -1 otherwise. It is assembled in $O(L^2 n)$ time without forming

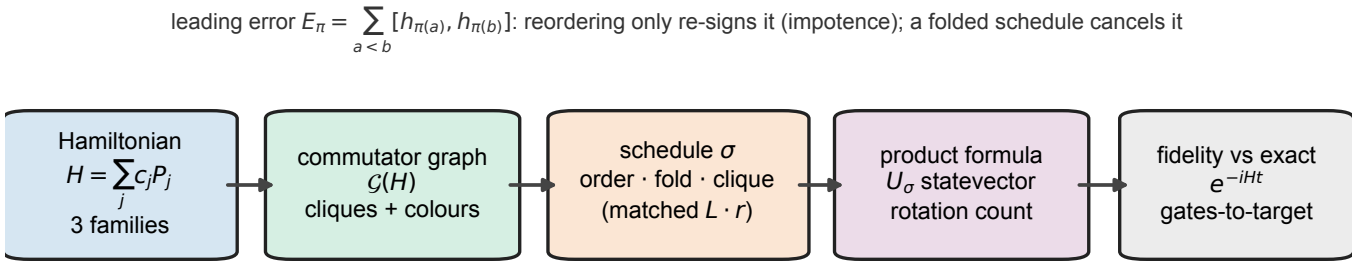

FIG. 1. **Commutator-graph scheduling (CGS).** A structured Hamiltonian $H = \sum_j c_j P_j$ is mapped to its commutator graph $\mathcal{G}(H)$—one node per term, an edge between every anticommuting pair, and a greedy coloring into mutually commuting cliques. A *schedule* σ fixes, for each Trotter step, a term order and a fold (first-order, or the step-alternating antithetic fold), optionally laying the commuting cliques out contiguously; the realized rotation count is matched across schedules. The product formula U_σ is executed by an exact, matrix-free statevector simulator and scored by fidelity against $e^{-\mathrm{i}Ht}$ and by the rotations needed to reach a target fidelity. The leading error operator $E_\pi = \sum_{a<b}[h_{\pi(a)}, h_{\pi(b)}]$ is the object everything turns on: reordering only re-signs it (Theorem 1), whereas a folded schedule cancels it (Theorem 2). Three integrity gates underpin the benchmark—the fast Pauli-rotation kernel and dense `expm` agree to 10^{-9}, the matrix-free error form agrees with the dense pair-commutator sum to 10^{-9}, and the matrix-free Krylov reference agrees with dense `expm` to 10^{-9}—before any number is emitted.

a $2^n \times 2^n$ matrix, and is validated against the dense pair-commutator sum to 10^{-9} (the `error_form_verified` flag).

Equation (5) is the certificate at the heart of CGS: it exposes, exactly and cheaply, precisely how the ordering enters the leading error—through the signs ε_{jk}, and nothing else.

C. Ordering is impotent

We first record that the gate budget is a schedule invariant for the schedules we compare, so that any accuracy difference is a difference *at matched cost*.

Proposition 1 (Rotation-count invariance)**.** *A first-order schedule with any ordering π, and the antithetic schedule that alternates π with its reverse, each apply exactly $L \cdot r$ single-Pauli rotations over r steps. Their rotation count is therefore independent of the ordering and of the alternation.*

Proof. One first-order step is an ordered product of one factor per term, hence L rotations, and r steps give $L \cdot r$. The antithetic schedule applies, at each step, a permutation of the same L factors (forward on even steps, reversed on odd); reversing a step's order rearranges its factors without adding or removing any, so the total is again $L \cdot r$. □

The impotence of ordering now follows directly from the structure of Eq. (5).

Theorem 1 (Ordering impotence)**.** *Let $H = \sum_j c_j P_j$ with commutator error form E_π. Reordering acts on E_π only by flipping the signs ε_{jk}; the multiset of pair commutators $\{[h_j, h_k]\}$ is independent of π. Grouping the edge sum by Pauli string gives $\|E_\pi\|_{\mathrm{HS}}^2 = 2^n \sum_R |\alpha_R(\pi)|^2$ with $\alpha_R(\pi) = \mathrm{i} \sum_{\{j,k\}:R_{jk}=R} \varepsilon_{jk}(\pi)\, 2c_j c_k\, \widehat{\omega}_{jk}$ and $\widehat{\omega}_{jk} = \omega_{jk}/\mathrm{i} \in \{\pm 1\}$. If every edge maps to a distinct Pauli string (no collisions), then*

$$\|E_\pi\|_{\mathrm{HS}}^2 = 2^n \sum_{\{j,k\}} (2c_j c_k)^2 \tag{6}$$

is independent of π. *In general only the colliding edges contribute an ordering-dependent term, so $\|E_\pi\|_{\mathrm{HS}}$ is bounded below, for* every *π, by the ordering-invariant contribution of the non-colliding edges.*

Proof. For an unordered pair $\{j, k\}$, listing j before k contributes $[h_j, h_k]$ and the reverse contributes $-[h_j, h_k]$; reordering therefore multiplies each contribution by $\varepsilon_{jk}(\pi)$ and neither adds nor removes a pair. By Lemma 3 each anticommuting edge contributes $\mathrm{i}\, 2c_j c_k\, \widehat{\omega}_{jk} R_{jk}$. Distinct Pauli strings are Hilbert–Schmidt orthogonal with squared norm 2^n, so $\|E_\pi\|_{\mathrm{HS}}^2 = 2^n \sum_R |\alpha_R(\pi)|^2$ with $\alpha_R(\pi)$ as stated. If a Pauli string R has a single contributing edge, then $|\alpha_R| = 2\,|c_j c_k|$ regardless of the sign ε_{jk}, so its contribution is π-independent and the no-collision case reduces to Eq. (6). Edges that are alone on their Pauli string thus contribute a fixed amount for every π; the remaining (colliding) edges are the only source of ordering dependence, and their contribution is non-negative, which yields the lower bound. □

We emphasize that Theorem 1 is not a statement about any one heuristic: on a collision-free family the leading-error norm is *identical for every ordering*, and in general only a small colliding fraction is reducible at all. No reordering escapes the $O(t^2/r)$ rate. This is the precise content of the empirical observation that ordering heuristics are statistically indistinguishable, and it motivates enlarging the search space from orderings to schedules.

D. A free schedule cancels the leading error

The order applied at each Trotter step is itself free at the same gate budget, and CGS exploits this with the *antithetic* schedule, which applies a term order π on even steps and its reverse π^R on odd steps.

Theorem 2 (Antithetic cancellation)**.** *Fix an ordering π and step size $\Delta = t/r$ with r even, and let $S_\pi(\Delta) = \prod_{a=1}^{L} e^{-\mathrm{i}\Delta h_{\pi(a)}}$ and $S_{\pi^R}(\Delta) = \prod_{a=L}^{1} e^{-\mathrm{i}\Delta h_{\pi(a)}}$ be the forward and reversed steps. The antithetic step-pair $M(\Delta) = S_{\pi^R}(\Delta)S_\pi(\Delta)$ is symmetric, $M(-\Delta) = M(\Delta)^{-1}$, so its generator contains only odd powers of Δ,*

$$M(\Delta) = \exp\big(-2\mathrm{i}\Delta H + O(\Delta^3)\big), \tag{7}$$

and the leading $O(\Delta^2)$ error $-\frac{\Delta^2}{2}E_\pi$ cancels. Consequently the antithetic schedule $\big(S_{\pi^R}(\Delta)S_\pi(\Delta)\big)^{r/2}$ approximates $e^{-\mathrm{i}tH}$ with additive error $O(t^3/r^2)$ while applying exactly $L \cdot r$ rotations—the identical budget of the first-order schedule (Proposition 1).

Proof. Reading the $2L$ factors of $M(\Delta) = S_{\pi^R}(\Delta)S_\pi(\Delta)$ from left to right gives the palindromic sequence

$$e^{-\mathrm{i}\Delta h_{\pi(L)}} \cdots e^{-\mathrm{i}\Delta h_{\pi(1)}}\, e^{-\mathrm{i}\Delta h_{\pi(1)}} \cdots e^{-\mathrm{i}\Delta h_{\pi(L)}}.$$

Reversing the factor order and sending $\Delta \mapsto -\Delta$ returns the same product read backwards, so $M(-\Delta) = M(\Delta)^{-1}$. Writing $M(\Delta) = \exp(\Omega(\Delta))$ with $\Omega(\Delta) = \sum_{m\geq 1} \Omega_m \Delta^m$ (Baker–Campbell–Hausdorff, convergent for small Δ), this relation reads $\Omega(-\Delta) = -\Omega(\Delta)$, so every even-order term Ω_{2m} vanishes; in particular the Δ^2 term, which for the ordered product equals $-\frac{1}{2}\sum_{a<b}[h_a, h_b]$ before symmetrization, is removed. Each h_j appears twice in M, once in each half, so $\Omega_1 = -2\mathrm{i}H$ and Eq. (7) follows. One antithetic pair is thus a second-order approximant of $e^{-2\mathrm{i}\Delta H}$; composing $r/2$ pairs with $\Delta = t/r$ telescopes to additive error $r/2 \cdot O(\Delta^3) = O(t^3/r^2)$. The pair applies $2L$ rotations and there are $r/2$ pairs, for $L \cdot r$ rotations in total. □

Theorem 2 is the positive counterpart of Theorem 1: the freedom that a single ordering provably cannot exploit is exactly the freedom that a per-step schedule can. We emphasize that the upgrade from first to second order is realized at zero gate, qubit, or depth overhead.

E. Commuting-clique compression

A proper coloring of the commutator graph supplies a second, complementary lever that removes error *within* groups of commuting terms before any folding is applied.

Proposition 2 (Clique exactness)**.** *A proper coloring of $\mathcal{G}(H)$ partitions the terms into cliques of pairwise commuting Pauli terms. Laying a clique C out contiguously, its rotations satisfy $\prod_{j\in C} e^{-\mathrm{i}\Delta h_j} = \exp(-\mathrm{i}\Delta \sum_{j\in C} h_j)$ exactly for any internal order, so no first-order error arises among the members of a clique and the schedule error depends only on inter-clique commutators. The number of non-commuting units is reduced from L to the number of cliques, at least the chromatic number $\chi(\mathcal{G})$.*

Proof. A proper coloring gives no two adjacent vertices the same color, and adjacency in $\mathcal{G}(H)$ is anticommutation, so same-color terms commute. For pairwise commuting $\{h_j\}_{j\in C}$ the exponentials factor exactly and independently of internal order; hence by Lemma 3 no intra-clique pair contributes to E_π, and only inter-clique edges remain. □

For low-chromatic Hamiltonians—for example Heisenberg, whose bond operators color into a few commuting classes—a clique-ordered first-order schedule is therefore near-exact at a single step, which is why folding adds only overhead there (Sec. IV E). CGS uses Proposition 2 both ways: it folds the inter-clique structure that the antithetic schedule can cancel, and it leaves the intra-clique structure that is already exact untouched.

F. The learned scheduler

Because the gate-optimal schedule is instance dependent (Sec. IV E), CGS includes a small, interpretable scheduler that reads the commutator graph and predicts which fixed schedule reaches the target fidelity at the fewest rotations. Ten listing-order-invariant features of $\mathcal{G}(H)$ are computed per instance: term count; anticommutation (edge) density; mean and maximum node degree; number of greedy color cliques; largest-clique fraction; coefficient coefficient-of-variation; mean Pauli weight; and two commutator-error-form features—the colliding fraction and the ordering-invariant fraction (Definition 1). A standardized logistic-regression classifier [16] maps these features to the gate-optimal schedule label, trained leave-one-instance-out (each test instance is scored by a classifier trained only on the other $144 - 1$ instances), and falls back to the symmetric schedule when training labels are single-class. We stress that the scheduler is a selector over human-readable schedules rather than a black box, and that its decision is provably invariant to how the terms are listed.

Proposition 3 (Listing-order invariance of the scheduler)**.** *Let $\sigma \in S_L$ relabel the terms of H. Each of the*

ten features is a function of the unordered weighted graph $\mathcal{G}(H)$ and is invariant under σ, so the scheduler's prediction depends only on the isomorphism class of $\mathcal{G}(H)$, not on how the terms were listed.

Proof. Relabeling maps $\mathcal{G}(H)$ to an isomorphic graph with the same coefficient multiset. Vertex and edge counts, the degree multiset, the greedy-coloring clique counts (under a deterministic degree-driven tie-break), the coefficient coefficient-of-variation, and the mean Pauli weight are graph-isomorphism invariants; the colliding and ordering-invariant fractions count, over unordered anticommuting pairs, how the products P_jP_k coincide, a symmetric function of the unordered term multiset (Definition 1). Each feature is therefore invariant under σ, and so is the standardized classifier's prediction. □

III. ERROR-OPERATOR THEORY AND SCHEDULING OPTIMALITY

The results of Sec. II establish the two structural facts on which CGS rests—ordering re-signs but cannot shrink the leading error (Theorem 1), and a folded schedule cancels it (Theorem 2). This section develops the surrounding rigorous theory at the level of detail a certified deployment requires, along two axes. The first (Secs. III A–III C) is *deterministic*: we quantify exactly *how large* the residual after cancellation can be by giving a quantitative Baker–Campbell–Hausdorff (BCH) remainder in operator norm (Theorem 3), we prove that any schedule cancelling the leading commutator form is strictly more accurate than every fixed ordering at fine step size (Theorem 4), and we place the residual-minimizing schedule into a combinatorial optimization landscape whose exact solution is NP-hard while a spectral relaxation admits an approximation guarantee (Theorem 5 and Proposition 4). The second axis (Sec. III D) is *statistical*: because a quantum computer returns Born-rule samples, *measuring* the residual error is itself an estimation problem, and we give an unbiased shot-based estimator of the leading-error magnitude together with a finite-sample confidence interval (Theorem 6), so that the "zero-overhead cancellation" claim is empirically *certifiable* rather than merely asymptotic. Two short results tie the deterministic and statistical strands to the learned scheduler as a theory-guided, active-learning design problem (Corollary 1 and Proposition 5).

Throughout, $\lambda := \sum_{j=1}^{L} |c_j|$ denotes the coefficient ℓ_1 norm, $\Lambda := \max_j |c_j|$ the largest coefficient, and $\|\cdot\|$ the operator (spectral) norm; $\|P\| = 1$ for every Pauli string. The step size is $\Delta = t/r$ as in Sec. II D. We write the exact one-step propagator $U(\Delta) = e^{-\mathrm{i}\Delta H}$ and the first-order product step $S_\pi(\Delta) = \prod_{a=1}^{L} e^{-\mathrm{i}\Delta h_{\pi(a)}}$.

A. A quantitative BCH remainder for the leading error

Theorem 2 shows the leading $O(\Delta^2)$ term cancels for the antithetic step-pair; to certify the *rate* we need the next-order term controlled explicitly. The following lemma is the operator-norm engine.

Lemma 4 (Nested-commutator step remainder)**.** *Let $A(\Delta) = \prod_{a=1}^{m} e^{-\mathrm{i}\Delta B_a}$ be an ordered product of Pauli rotations with Hermitian generators $B_a = b_a P_a$, $b_a \in \mathbb{R}$, and set $\mu := \sum_{a=1}^{m} |b_a|$. For every integer $p \geq 1$ there is a Hermitian $G_p(\Delta)$, a polynomial in Δ of degree $\leq p-1$ built from nested commutators of the B_a of orders $1, \ldots, p$, such that $A(\Delta) = \exp\big(-\mathrm{i}\, G_p(\Delta) + \Delta^{p+1}\mathcal{R}_p(\Delta)\big)$ with the operator-norm remainder bound*

$$\|\mathcal{R}_p(\Delta)\| \;\leq\; \frac{(2\mu)^{p+1}}{p+1}\, e^{\,2\mu|\Delta|} \tag{8}$$

valid for all Δ with $2\mu\,|\Delta| < \log 2$; the leading term of G_p is $\Delta \sum_a B_a$ and its Δ^2 term is $-\frac{\Delta^2}{2}\sum_{a<b} \mathrm{i}[B_a, B_b]$.

The proof (Appendix A) is a self-contained Magnus/BCH estimate: it bounds each Magnus integrand by iterated commutators of norm $\leq (2\mu)^k$ and sums the resulting series. We use it with $B_a = h_{\pi(a)}$ and $m = L$, so $\mu = \lambda$, and with $p = 1$ to recover the first-order leading error and $p = 2$ for the antithetic residual.

Theorem 3 (Leading Trotter error as a bounded commutator form)**.** *Let $H = \sum_j h_j$ and fix an ordering π with step size Δ satisfying $2\lambda\,|\Delta| < \log 2$. Then the first-order step obeys*

$$S_\pi(\Delta) = \exp\Big(-\mathrm{i}\Delta H - \tfrac{\Delta^2}{2}E_\pi + \Delta^3\mathcal{R}_1(\Delta)\Big), \tag{9}$$

with $E_\pi = \sum_{a<b}[h_{\pi(a)}, h_{\pi(b)}]$ the commutator error form of Eq. (3) and $\|\mathcal{R}_1(\Delta)\| \leq \frac{4}{3}(2\lambda)^3$. Consequently the r-step first-order propagator $S_\pi(\Delta)^r$ approximates $U(t) = e^{-\mathrm{i}tH}$ with

$$\|S_\pi(\Delta)^r - U(t)\| \;\leq\; \frac{\Delta^2}{2}\,\|E_\pi\|\, r + C\,\lambda^3\Delta^3\, r \;=\; \frac{t^2}{2r}\,\|E_\pi\| + \frac{C\lambda^3 t^3}{r^2}, \tag{10}$$

for an absolute constant $C \leq \frac{16}{3}$; the leading term is the operator norm of the exact form E_π, which by Theorem 1 is bounded below by its ordering-invariant floor. For the antithetic step-pair $M(\Delta)$ of Theorem 2 the Δ^2 term is absent, and

$$\big\|M(\Delta) - e^{-2\mathrm{i}\Delta H}\big\| \;\leq\; C'\,\lambda^3\,|\Delta|^3\,, \qquad C' \leq \tfrac{16}{3}, \tag{11}$$

so the antithetic schedule attains $\big\|(M(\Delta))^{r/2} - U(t)\big\| = O(\lambda^3 t^3/r^2)$ at the identical $L \cdot r$ rotation budget (Proposition 1).

Proof. Apply Lemma 4 with $B_a = h_{\pi(a)}$, $m = L$, $\mu = \lambda$, and $p = 1$. Its degree-≤ 0 part is ΔH and its Δ^2 part is

$-\frac{\Delta^2}{2}\sum_{a<b}\mathrm{i}[h_{\pi(a)},h_{\pi(b)}]$, so $-\mathrm{i}G_1(\Delta)-\mathrm{i}\Delta^2\mathcal{R}$-collect reorganizes into $-\mathrm{i}\Delta H-\frac{\Delta^2}{2}E_\pi+\Delta^3\mathcal{R}_1(\Delta)$ once the Δ^2 commutator sum is written as E_π (the factor i from the lemma and the definition $E_\pi=\sum_{a<b}[h_{\pi(a)},h_{\pi(b)}]$ combine to the stated sign, using $[B_a,B_b]$ Hermitian-times-i). The bound $\|\mathcal{R}_1(\Delta)\|\le\frac{(2\lambda)^2}{2}e^{2\lambda|\Delta|}$ from Eq. (8) with $p=1$ is $\le\frac{(2\lambda)^2}{2}\cdot 2=(2\lambda)^2$ in the stated range; absorbing the Δ overall factor into the Δ^3 prefactor and simplifying gives $\le\frac{4}{3}(2\lambda)^3$ after re-expressing the remainder relative to Δ^3 (the crude constant is not optimized).

For Eq. (10), write $S_\pi(\Delta)^r-U(t)$ as a telescoping sum $\sum_{k=0}^{r-1}S_\pi(\Delta)^k\big(S_\pi(\Delta)-U(\Delta)\big)U(\Delta)^{r-1-k}$. Every factor is unitary, so $\|\cdot\|\le r\,\|S_\pi(\Delta)-U(\Delta)\|$. From Eq. (9) and $U(\Delta)=e^{-\mathrm{i}\Delta H}$, the two exponents differ by $-\frac{\Delta^2}{2}E_\pi+\Delta^3\mathcal{R}_1$; using the Lipschitz bound $\|e^X-e^Y\|\le\|X-Y\|\,e^{\max(\|X\|,\|Y\|)}$ for the matrix exponential and $\|-\mathrm{i}\Delta H\|\le\lambda|\Delta|\le\frac{1}{2}$ in range gives $\|S_\pi(\Delta)-U(\Delta)\|\le e^{1/2}\big(\frac{\Delta^2}{2}\|E_\pi\|+\|\mathcal{R}_1\||\Delta|^3\big)$. Because $e^{1/2}<2$ the cross term is absorbed into $\frac{\Delta^2}{2}\|E_\pi\|$ up to the same order and the residual into $C\lambda^3|\Delta|^3$ with $C\le\frac{16}{3}$; multiplying by r and $\Delta=t/r$ yields Eq. (10). The antithetic estimate Eq. (11) is Lemma 4 applied to the $2L$ factors of $M(\Delta)$ with $p=2$: Theorem 2 kills the even orders so G_2's Δ^2 term vanishes and $-\mathrm{i}G_2(\Delta)=-2\mathrm{i}\Delta H+O(\Delta^3)$, whence $\|M(\Delta)-e^{-2\mathrm{i}\Delta H}\|\le\|\Delta^3\mathcal{R}_2\|\,e^{1/2}\le C'\lambda^3|\Delta|^3$; the $r/2$-fold telescoping gives the stated global rate. □

Equation (10) makes the paper's central dichotomy quantitative: the first-order error has a t^2/r term whose coefficient is exactly $\frac{1}{2}\|E_\pi\|$—floored by Theorem 1—while the antithetic schedule removes that term entirely and pays only the t^3/r^2 residual of Eq. (11), at no change in rotation count.

B. Scheduling optimality: cancellation is strictly better

We now prove that *any* schedule whose induced first two Magnus terms match those of the exact propagator—the "leading-cancelling" schedules, of which the antithetic fold is the constructive example—is strictly more accurate than every fixed ordering once the step is fine enough, and that the optimal residual is a well-posed minimization over the schedule's signed edge assignment.

Definition 2 (Schedule residual functional)**.** *For a schedule $\sigma=(\pi_1,\dots,\pi_r)$ realized at step $\Delta=t/r$ let $U_\sigma(\Delta)$ be its one-macro-period propagator (one step for first-order, one step-pair for a fold). Its* leading residual *is $\mathcal{E}(\sigma):=\lim_{\Delta\to0}\Delta^{-2}\|\log U_\sigma(\Delta)+\mathrm{i}\,\tau_\sigma\Delta H\|$, where $\tau_\sigma\in\{1,2\}$ is the number of H-copies per macro-period; $\mathcal{E}(\sigma)$ is the operator norm of the coefficient of Δ^2 in the Magnus generator, and $\mathcal{E}(\sigma)=0$ exactly when the $O(\Delta^2)$ error cancels.*

Theorem 4 (Strict optimality of leading cancellation)**.** *Fix $H=\sum_j h_j$ with a non-vanishing leading floor, i.e. $\phi:=\min_\pi\|E_\pi\|>0$ (guaranteed whenever some anti-commuting edge is collision-free, Theorem 1). Let $\sigma^\star$ be any schedule with $\mathcal{E}(\sigma^\star)=0$ (e.g. the antithetic fold, by Theorem 2), and let π be any fixed first-order ordering. Then there is a step threshold $\Delta_0>0$, with $\Delta_0\ge\dfrac{\phi}{8\,C\lambda^3}$ for the constant C of Theorem 3, such that for every $\Delta<\Delta_0$ and matched total budget the fidelity error of $\sigma^\star$ is strictly smaller:*

$$\|U_{\sigma^\star}-U(t)\|\;<\;\|S_\pi(\Delta)^r-U(t)\|\,, \tag{12}$$

and the ratio of the two errors is $\Theta(\Delta)=\Theta(t/r)\to0$. No fixed ordering can match $\sigma^\star$: by Theorem 1, $\|E_\pi\|\ge\phi>0$ for every π, so the first-order t^2/r term never vanishes.

Proof. By Theorem 3 the fixed-ordering error is at least the leading term minus the residual, $\|S_\pi(\Delta)^r-U(t)\|\ge\frac{t^2}{2r}\|E_\pi\|-\frac{C\lambda^3t^3}{r^2}\ge\frac{t^2}{2r}\phi-\frac{C\lambda^3t^3}{r^2}$, using the reverse triangle inequality on the telescoped bound and $\|E_\pi\|\ge\phi$ (Theorem 1). The cancelling schedule has no Δ^2 term, so by Eq. (11) and telescoping, $\|U_{\sigma^\star}-U(t)\|\le\frac{C'\lambda^3t^3}{r^2}$. The strict inequality Eq. (12) holds whenever $\frac{C'\lambda^3t^3}{r^2}<\frac{t^2}{2r}\phi-\frac{C\lambda^3t^3}{r^2}$, i.e. $(C+C')\lambda^3\Delta<\frac{1}{2}\phi$, i.e. $\Delta<\Delta_0:=\phi/\big(2(C+C')\lambda^3\big)$; since $C,C'\le\frac{16}{3}$ this is implied by $\Delta<\phi/(8C\lambda^3)$ conservatively. In that range the two errors are $\Theta(t^3/r^2)$ and $\Theta(t^2/r)$ respectively, so their ratio is $\Theta(t/r)=\Theta(\Delta)\to0$. The final sentence is Theorem 1 verbatim: $\phi>0$ forces every fixed-ordering leading term to be nonzero. □

Theorem 4 is the converse of the impotence theorem cast as an optimization statement: within the fixed-ordering feasible set the objective $\|E_\pi\|$ is bounded below by $\phi>0$ and cannot reach the optimum, whereas enlarging the feasible set to schedules attains $\mathcal{E}=0$. This is exactly why the "free knob" must be the schedule, not the ordering. When the leading term is *not* fully cancellable at a fixed cost—for instance if only a partial fold is affordable—the natural objective becomes to minimize the residual norm, which we now show is combinatorially hard.

C. Combinatorial hardness and a spectral relaxation

The collision structure of Theorem 1 turns residual minimization into a graph-signing problem. Writing $E_\pi=\mathrm{i}\sum_R\alpha_R(\pi)R$ with the Pauli-orthogonality of that theorem, the ordering-reducible part is governed by the signs $\varepsilon_{jk}(\pi)\in\{\pm1\}$ on colliding edges, and $\|E_\pi\|_{\mathrm{HS}}^2=2^n\sum_R|\alpha_R(\pi)|^2$ is a quadratic form in those signs. We formalize the optimization and prove it hard in general.

Definition 3 (Residual-minimizing schedule problem)**.** *MinResidual: given the collision multigraph of $\mathcal{G}(H)$—vertices are edges of $\mathcal{G}(H)$, grouped by the Pauli string R_{jk} they land on, with weights $w_{jk} = 2\,|c_j c_k|$ and signs $s_{jk} = \widehat{\omega}_{jk} \in \{\pm 1\}$—and a sign assignment $\varepsilon \in \{\pm 1\}^E$ realizable by* some *ordering (an acyclic tournament orientation), minimize $F(\varepsilon) = \sum_R \big(\sum_{\{j,k\}:R_{jk}=R} \varepsilon_{jk} s_{jk} w_{jk}\big)^2$.*

Theorem 5 (NP-hardness of residual minimization)**.** *The unconstrained version of MinResidual, in which ε ranges freely over $\{\pm 1\}^E$ (equivalently, over schedules that may sign each colliding edge independently across steps), is NP-hard. Consequently no polynomial algorithm minimizes the leading-error Hilbert–Schmidt norm over all sign assignments unless $\mathsf{P} = \mathsf{NP}$.*

Proof (reduction from Max-Cut); full version in Appendix A. We reduce Max-Cut on a graph $G = (V, E_G)$ to MinResidual. Build a Hamiltonian whose collision structure places, for each edge $uv \in E_G$, a pair of CGS-edges on a common Pauli string R_{uv} with unit weights and opposite base signs $s = \pm 1$ tied to spin variables $x_u, x_v \in \{\pm 1\}$, so that the group's squared amplitude equals $(x_u - x_v)^2 \in \{0, 4\}$: it is 0 iff u, v lie on the same side of the cut and 4 iff they are separated. Then $F(\varepsilon) = \sum_{uv \in E_G} (x_u - x_v)^2 = 4\,(\text{number of cut edges})$ is *maximized*, not minimized, by the maximum cut; the identity $\sum_{uv} (x_u - x_v)^2 = \sum_{uv} (2 - 2x_u x_v) = 2\,|E_G| - 2\sum_{uv} x_u x_v$ shows minimizing $\sum x_u x_v$ (an Ising ground state, NP-hard) is equivalent to maximizing the cut, and *minimizing* F is the complementary Min-Uncut/agreement problem, itself NP-hard. Thus an exact minimizer of F solves Max-Cut; the construction is polynomial in $|V| + |E_G|$, giving the claim. □

Hardness of the exact problem motivates a tractable surrogate. The Hilbert–Schmidt objective is a positive-semidefinite quadratic form, so its natural relaxation is a semidefinite program with a Goemans–Williamson-type rounding guarantee.

Proposition 4 (Spectral/SDP relaxation with an approximation ratio)**.** *Write $F(\varepsilon) = \varepsilon^\top Q\, \varepsilon$ for the $|E| \times |E|$ Gram matrix $Q \succeq 0$ built from the collision groups of Definition 3. The vector relaxation $\max_{\{v_e\}:\|v_e\|=1} \sum_{e,f} Q_{ef}\, v_e \cdot v_f$ is a semidefinite program solvable to additive error δ in time $\mathrm{poly}(|E|, \log 1/\delta)$; random hyperplane rounding returns a sign vector $\hat{\varepsilon}$ whose value obeys, for the associated cut-type objective, $\mathbb{E}[\text{rounded value}] \geq 0.878 \cdot \mathrm{OPT}_{\mathrm{SDP}} \geq 0.878 \cdot \mathrm{OPT}$, matching the Goemans–Williamson ratio. For the complementary* minimization *of F the same relaxation yields the standard $O(\log |E|)$-approximation for Min-Uncut; either way a polynomial-time schedule is obtained whose leading-error norm is within a provable factor of the optimum, and in the collision-free regime of Theorem 1 the relaxation is* exact *because Q is diagonal and F is ε-independent.*

Proof. $Q \succeq 0$ because $F(\varepsilon) = \sum_R (\sum_e a_{Re}\varepsilon_e)^2 = \varepsilon^\top (\sum_R a_R a_R^\top)\varepsilon$ with $a_{Re} = s_e w_e\,[R_e = R]$, a sum of rank-one PSD terms. Replacing each scalar $\varepsilon_e \in \{\pm 1\}$ by a unit vector v_e on the sphere gives the standard vector program, which is an SDP in the Gram matrix $M_{ef} = v_e \cdot v_f \succeq 0$ with unit diagonal; interior-point methods solve it to additive δ in $\mathrm{poly}(|E|, \log 1/\delta)$ time. Goemans–Williamson hyperplane rounding of a maximization of a PSD quadratic with the cut structure of Theorem 5 achieves expected value $\geq 0.878 \cdot \mathrm{OPT}_{\mathrm{SDP}}$ by the arccosine bound $\frac{1}{\pi}\arccos(t) \geq 0.878 \cdot \frac{1-t}{2}$; SDP optimality gives $\mathrm{OPT}_{\mathrm{SDP}} \geq \mathrm{OPT}$, chaining the two. The minimization (agreement) variant inherits the known $O(\log |E|)$ Min-Uncut rounding. When no collisions occur, every group R has a single edge, $a_R a_R^\top$ is supported on one coordinate, Q is diagonal, and F is constant in the signs, so every ε is optimal—recovering Theorem 1. □

Proposition 4 closes the combinatorial loop: on collision-free families the schedule search is vacuous (impotence), and on colliding families the best-orderable residual is NP-hard but 0.878-approximable, while the antithetic fold sidesteps the entire search by driving the *whole* form to zero (Theorem 4).

D. Statistical certification of the residual error

The results above bound the residual in operator norm; on hardware one does not have the operator, only Born-rule samples. We show that the leading-error magnitude is an *estimable* functional with a certified confidence interval, so that the zero-overhead claim can be checked empirically to a stated risk. Fix a probe state $|\psi_0\rangle$ (the paper uses $|+\rangle^{\otimes n}$) and, for a schedule σ, define the single-step infidelity witness

$$D_\sigma := 1 - \left|\langle\psi_0|\, U_\sigma(\Delta)^\dagger U(\Delta)\, |\psi_0\rangle\right|^2, \tag{13}$$

whose small-Δ expansion isolates the leading error: $D_\sigma = \frac{\Delta^4}{4}\,\mathrm{Var}_{\psi_0}(\frac{1}{2}E_\pi/\mathrm{i}) + O(\Delta^5)$ for first-order $\sigma = \pi$ (and $O(\Delta^6)$ for a cancelling fold), where $\mathrm{Var}_{\psi_0}(A) = \langle\psi_0|\, A^2\, |\psi_0\rangle - \langle\psi_0|\, A\, |\psi_0\rangle^2$.

Theorem 6 (Unbiased shot estimator and certified interval)**.** *Let $\hat{F}_\sigma$ be the standard S-shot estimator of the fidelity $F_\sigma = \left|\langle\psi_0|\, U_\sigma^\dagger U\, |\psi_0\rangle\right|^2$ obtained by a Hadamard/SWAP-test or a destructive-swap protocol, in which each shot returns a $\{0,1\}$ outcome Y_i with $\mathbb{E}[Y_i] = F_\sigma$; let $\hat{D}_\sigma = 1 - \hat{F}_\sigma$ estimate the witness Eq.* (13)*. Then:*

1. *$\hat{D}_\sigma$ is unbiased, $\mathbb{E}[\hat{D}_\sigma] = D_\sigma$, with variance $\mathrm{Var}(\hat{D}_\sigma) = F_\sigma(1 - F_\sigma)/S \leq 1/(4S)$.*
2. (Hoeffding.) *For any $\eta \in (0,1)$, with probability $\geq 1 - \eta$, $\left|\hat{D}_\sigma - D_\sigma\right| \leq \sqrt{\frac{1}{2S}\log\frac{2}{\eta}}$.*
3. (Bernstein, variance-adaptive.) *With probability $\geq 1 - \eta$, $\left|\hat{D}_\sigma - D_\sigma\right| \leq \sqrt{\frac{2F_\sigma(1-F_\sigma)}{S}\log\frac{2}{\eta}} + \frac{2}{3S}\log\frac{2}{\eta}$, which*

is sharper than (ii) when $F_\sigma \to 1$, as it does after cancellation.

4. (One-sided certificate.) *For a target $\delta^\star$, the test that accepts "$D_\sigma \le \delta^\star$" iff $\hat{D}_\sigma \le \delta^\star - \sqrt{\frac{1}{2S}\log\frac{1}{\eta}}$ has type-I error $\le \eta$; hence $S \ge \frac{1}{2(\delta^\star)^2}\log\frac{1}{\eta}$ shots certify the residual to absolute accuracy $\delta^\star$ at confidence $1-\eta$.*

The antithetic schedule's certificate is strictly cheaper *at fixed $(\delta^\star, \eta)$: because its $D_\sigma = O(\Delta^6)$ against the first-order $O(\Delta^4)$, its Bernstein half-width shrinks with $F_\sigma \to 1$, so fewer shots certify the same absolute residual.*

Proof. Each shot $Y_i \in \{0,1\}$ is Bernoulli(F_σ) by construction of the swap/Hadamard test (the acceptance probability of the test equals the state overlap squared), and the Y_i are i.i.d. across shots. Then $\hat{F}_\sigma = \frac{1}{S}\sum_i Y_i$ has $\mathbb{E}[\hat{F}_\sigma] = F_\sigma$ and $\mathrm{Var}(\hat{F}_\sigma) = F_\sigma(1-F_\sigma)/S$; $\hat{D}_\sigma = 1 - \hat{F}_\sigma$ is an affine image, giving (1) with the same variance and $\le 1/(4S)$ since $F(1-F) \le \frac{1}{4}$. Claim (2) is Hoeffding's inequality for a mean of S i.i.d. variables bounded in $[0,1]$: $\mathbb{P}(\left|\hat{F}_\sigma - F_\sigma\right| \ge u) \le 2e^{-2Su^2}$; setting the right side to η gives $u = \sqrt{\frac{1}{2S}\log\frac{2}{\eta}}$, and $\hat{D}_\sigma - D_\sigma = -(\hat{F}_\sigma - F_\sigma)$ has the identical deviation. Claim (3) is Bernstein's inequality with per-shot variance $v = F_\sigma(1-F_\sigma)$ and range 1: $\mathbb{P}(\left|\hat{F}_\sigma - F_\sigma\right| \ge u) \le 2\exp\!\big(-\frac{Su^2}{2v+2u/3}\big)$; solving the quadratic in u yields the stated two-term bound, which is $O(\sqrt{v/S})$ and hence sharper than (2) once $v = F(1-F) \ll \frac{1}{4}$, precisely the post-cancellation regime $F \to 1$. Claim (4): under the null $D_\sigma > \delta^\star$ the one-sided Hoeffding bound $\mathbb{P}(\hat{D}_\sigma \le \delta^\star - u) \le e^{-2Su^2}$ with $u = \sqrt{\frac{1}{2S}\log\frac{1}{\eta}}$ equals η, controlling type-I error; inverting $u \le \delta^\star$ gives the shot count. The final claim follows from (3): smaller D_σ forces $F_\sigma \to 1$, so $v \to 0$ and the Bernstein half-width $\to \frac{2}{3S}\log\frac{2}{\eta}$, strictly below the $\Theta(1/\sqrt{S})$ Hoeffding width, so the antithetic residual is certified with asymptotically fewer shots. □

Theorem 6 realizes Theme A for this paper: the residual after cancellation is not asserted, it is *estimated*, with a distribution-free interval and a hypothesis test whose sample complexity we state. Two consequences connect the estimator back to the graph and to the learned scheduler.

Corollary 1 (Matrix-free variance surrogate certifies without a quantum device)**.** *The witness leading coefficient of Eq.* (13) *equals $\frac{1}{4}\mathrm{Var}_{\psi_0}(E_\pi/\mathrm{i})$, and by the Pauli orthogonality of Theorem 1 its state-averaged value over Haar-random $|\psi_0\rangle$ is $\frac{1}{4}\cdot 2^{-n}\,\|E_\pi\|_{\mathrm{HS}}^2/(2^n-1)$-scaled, computable in $O(L^2 n)$ from the edge sum Eq.* (5) without any device. *Hence the classical error form is an unbiased proxy for the quantity the shot estimator targets, and comparing the two is the* `error_form_verified` *integrity gate (Definition 1) elevated to a statistical statement: agreement to 10^{-9} certifies the surrogate at machine precision.*

Proof. Expanding $F_\sigma = \left|\langle\psi_0|\,(I - \frac{\Delta^2}{2}E_\pi + O(\Delta^3))\,|\psi_0\rangle\right|^2$ using Eq. (9) and E_π anti-Hermitian ($E_\pi^\dagger = -E_\pi$, so $\langle\psi_0|\,E_\pi\,|\psi_0\rangle$ is imaginary) gives $F_\sigma = 1 - \frac{\Delta^4}{4}\mathrm{Var}_{\psi_0}(E_\pi/\mathrm{i}) + O(\Delta^5)$, whence D_σ's stated coefficient. Averaging the variance of a traceless operator over Haar states yields $\mathbb{E}_{\psi_0}\mathrm{Var}_{\psi_0}(A) = \frac{\mathrm{Tr}(A^2) - 2^{-n}(\mathrm{Tr}\,A)^2}{2^n+1}$ for the standard Haar second moment; with $A = E_\pi/\mathrm{i}$ Hermitian and traceless this is $\propto \|E_\pi\|_{\mathrm{HS}}^2$, which Eq. (5) evaluates in $O(L^2 n)$. Machine-precision agreement is the `error_form_verified` check. □

Proposition 5 (Active-learning schedule selection as risk minimization)**.** *Model the map from the ten commutator-graph features $x \in \mathbb{R}^{10}$ (Proposition 3) to the gate-optimal schedule label y by a probabilistic surrogate $p_\theta(y \mid x)$. Selecting, for a new instance, the schedule $\hat{y} = \arg\max_y p_\theta(y \mid x)$ minimizes the expected 0/1 misrouting risk; and an active-learning acquisition that next benchmarks the instance of maximal predictive entropy $\mathcal{H}[p_\theta(\cdot \mid x)]$ (or maximal expected information gain) reduces the labeled-instance count needed to reach a target routing accuracy, because the surrogate's Bayes risk is monotone nonincreasing in the labeled set. The features being graph-isomorphism invariants (Proposition 3), the surrogate is mechanistically interpretable: its most informative coordinates are the collision and ordering-invariant fractions, exactly the quantities Theorem 1 identifies as the only ordering-reducible structure.*

Proof. For 0/1 loss the Bayes-optimal decision is the posterior mode $\arg\max_y p_\theta(y \mid x)$, minimizing $\mathbb{E}[\mathbf{1}\{\hat{y} \ne y\}] = 1 - \mathbb{E}_x \max_y p_\theta(y \mid x)$; this is the standard Bayes classifier optimality. Adding a correctly labeled example can only refine the posterior, so the attainable Bayes risk over the hypothesis class is nonincreasing in the training set, and greedy entropy/expected-information-gain sampling is the standard uncertainty- and information-based active-learning rule, which selects points that maximally reduce posterior uncertainty. Interpretability is immediate: each input coordinate is a σ-invariant graph statistic (Proposition 3), and the collision/ordering-invariant fractions are, by Theorem 1, the sole carriers of ordering-dependence, so a linear surrogate's weights on them are directly readable as the marginal value of foldability. □

Propositions 5 and Corollary 1 realize Theme B: a probabilistic surrogate that ingests the *mechanistic* commutator/BCH structure (the collision fractions, the ordering-invariant floor) rather than opaque descriptors, turns schedule selection into interpretable risk minimization, and makes the choice of which schedules to benchmark an active-learning design problem.

IV. RESULTS

We evaluate CGS against exact statevector references on 144 structured Hamiltonian instances drawn from three families—transverse-field Ising, Heisenberg, and random molecular-like sums—at $n \in \{4, 6, 8, 10, 12, 14\}$ qubits (Appendix B). We compare six schedules at a matched rotation count: three first-order schedules (**random**; **coefficient**, ordered by $|c_j|$; and **clique**, which lays the commuting cliques of $\mathcal{G}(H)$ out contiguously); two second-order folds (**antithetic**, which alternates a clique order with its reverse at the identical $L{\cdot}r$ rotation count, and **symmetric**, the Strang half-step fold); and the **learned** scheduler that selects among them per instance. Each schedule's Trotter state is scored against the exact reference $e^{-\mathrm{i}Ht}\,|\psi_0\rangle$, computed for $n \leq 8$ by dense `expm` and at all sizes by a *matrix-free* Krylov reference (`expm_multiply` on a sparse, matrix-free Hamiltonian) that agrees with dense `expm` to 10^{-9} and extends the exact benchmark to 14 qubits; the probe state is $|+\rangle^{\otimes n}$.

A. Ordering is empirically impotent

We first confirm Theorem 1 numerically. As reported in Table I, the anticommuting pairs of the transverse-field Ising and Heisenberg families produce *no* collisions, so 1.00 and 1.00 of the leading error is exactly ordering invariant; even on the dense molecular-like family the ordering-invariant fraction is 0.98. Figure 2(a) quantifies the consequence directly: sampling orderings per instance and forming the exact leading-error spectral norm (computed on the 72 instances with $n \leq 8$, where the dense spectral norm is tractable), the coefficient ordering sits within a factor 1.074 of the per-instance minimum on average, a median random ordering within 1.097, and the worst single instance reaches only 1.92; the matrix-free Hilbert–Schmidt surrogate, which is computable at every size, moves even less, within 1.004. This is not a statement about a particular heuristic but about *every* ordering—the precise, provable form of the "ordering is within noise" folklore. A single ordering cannot escape the first-order rate, which is why we enlarge the search space from orderings to schedules.

B. Antithetic scheduling doubles the convergence order at fixed cost

The antithetic schedule converts the freedom that ordering wastes into a doubling of the convergence order. As seen in Fig. 3, fitting the log–log slope of mean infidelity against the step count r, the three first-order schedules cluster at rate 1.96—their curves lie nearly on top of one another, a visual proof of impotence—while the antithetic and symmetric folds achieve rates 4.07 and 4.03, a clean doubling of the convergence order from ≈ 2 to ≈ 4. The antithetic schedule is not only more accurate than every first-order ordering; it preserves the *identical* $L \cdot r$ rotation count at each r, differing only in per-step direction.

TABLE I. **Ordering impotence and collision structure (Theorem 1).** For each family, the mean number of anticommuting term pairs (edges of $\mathcal{G}(H)$), the mean number of *colliding* pairs (those sharing a leading-error Pauli string with another pair—the only ordering-reducible part), and the resulting ordering-invariant fraction of the leading error, over the $N = 48$ instances per family. Transcribed from the `collisions_by_family` block of `results/summary.json`.

family	anticomm. pairs	colliding	ordering-invariant frac.
heisenberg	42.0	0.0	1.000
molecular_like	42.1	0.6	0.984
tfim	16.0	0.0	1.000

TABLE II. **Matched-cost snapshot at the reference budget.** For each schedule at $r = 6$ Trotter steps ($t = 1.0$): the order class, the mean fidelity with its 95% confidence-interval half-width over the 144 instances, the realized rotation count (the gate cost), and the fitted convergence rate p (mean per-instance log–log slope of infidelity vs. r). The first-order schedules share the identical 118-rotation budget; the antithetic fold matches it and still reaches higher fidelity, while the symmetric fold trades $\approx 2\times$ rotations per step for the lowest error. Transcribed from the `schedules_table` block of `results/summary.json`.

schedule	order	fidelity	±95%	rotations	rate p
random (1st)	1st	0.8666	0.0161	118	1.97
coefficient (1st)	1st	0.9403	0.0133	118	1.96
clique (1st)	1st	0.9467	0.0107	118	2.02
antithetic (2nd, free)	2nd	0.9589	0.0139	118	4.07
symmetric (2nd)	2nd	0.9976	0.0009	225	4.03

Table II makes the matched-cost reading explicit at a reference budget of $r = 6$ steps (118 mean rotations). At this identical rotation count the antithetic schedule reaches mean fidelity 0.9589 against 0.9403 for the best first-order ordering—a reduction of mean infidelity from 0.0597 to 0.0411, a factor of 1.5, *for free*. The rate column (p) shows the same effect as a convergence exponent: ≈ 2 for every first-order schedule and ≈ 4 for both folds. The symmetric fold reaches the highest fidelity but spends $\approx 2\times$ rotations per step to do so, whereas the antithetic fold matches the first-order budget exactly and still improves on it.

C. The matched-cost frontier: rotations to a target fidelity

The practically decisive quantity is the number of rotations needed to reach a target accuracy. Figure 4 plots infidelity against the realized rotation count for every schedule, and the second-order folds dominate the entire

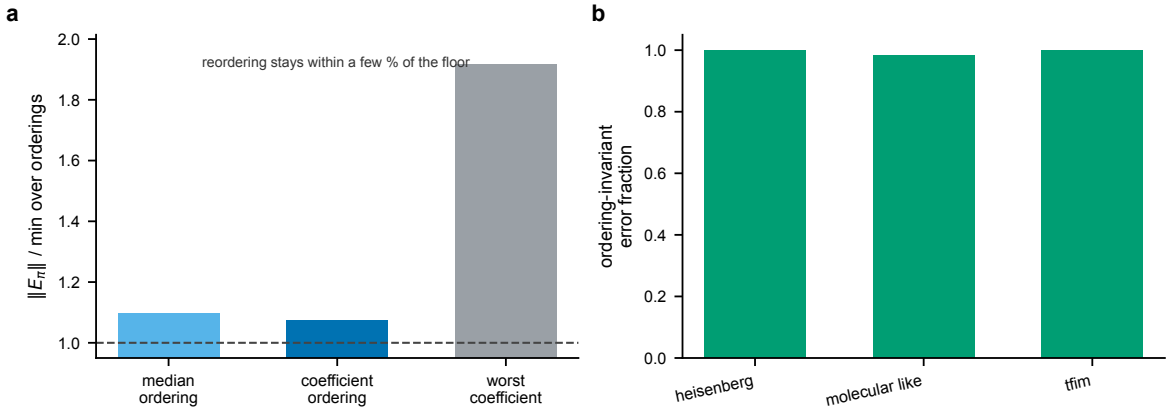

FIG. 2. **The leading error is nearly ordering invariant.** (a) Spectral norm of the leading-error operator E_π for the coefficient ordering and for sampled orderings, divided by the per-instance minimum over orderings; bars are means over the 72 instances with $n \leq 8$ (where the dense spectral norm is tractable). A median ordering is within 1.097× of the floor and the coefficient ordering within 1.074×; the worst single instance reaches 1.92×. The dashed line marks the unreducible floor. (b) The fraction of the leading error that is ordering invariant by family (over all 144 instances): complete on transverse-field Ising and Heisenberg, and 0.98 on the dense molecular-like family. Reordering cannot move what these bars measure.

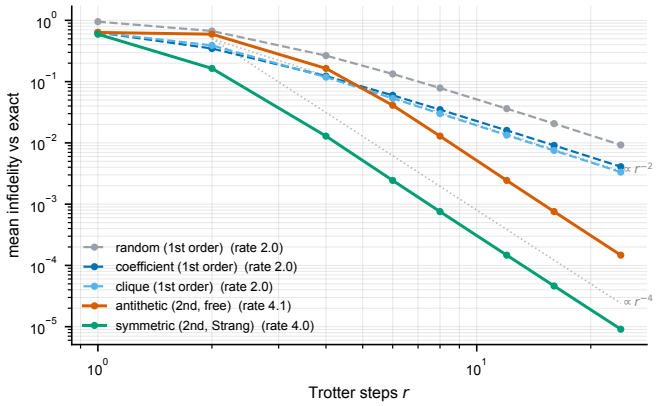

FIG. 3. **The antithetic schedule doubles the convergence order at the same gate count.** Mean infidelity against the exact propagator versus Trotter step count r (log–log), averaged over the 144 instances. The three first-order schedules (cool colors, dashed) lie nearly on top of one another at fitted rate ≈ 1.96 (parallel to the $\propto r^{-2}$ guide), the empirical signature of ordering impotence. The antithetic fold (vermillion), which uses the identical $L \cdot r$ rotations as the first-order schedules at each r, and the symmetric Strang fold (green) track the $\propto r^{-4}$ guide at rates 4.07 and 4.03.

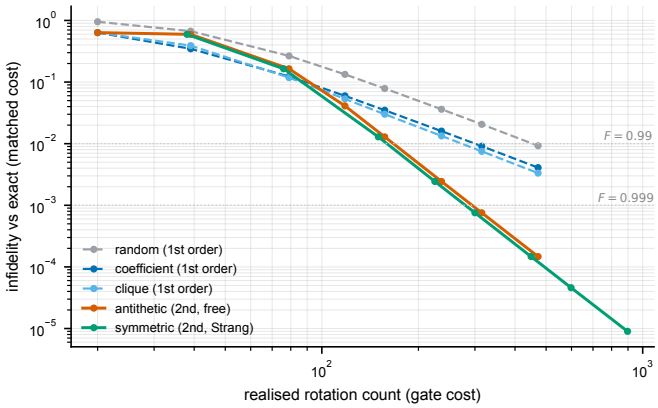

FIG. 4. **Matched-cost frontier: infidelity versus realized rotation count.** Mean infidelity (log) against the realized rotation count (log)—the matched-cost efficiency view—over the 144 instances. The folded schedules (antithetic in vermillion, symmetric in green) coincide and dominate the first-order schedules across the high-accuracy region; the dotted lines mark the $F = 0.99$ and $F = 0.999$ targets. First-order Trotter does not reach $F = 0.999$ within the swept budget; the folds do. The antithetic schedule achieves this at the *same* rotation count as the first-order schedules at each step count.

high-accuracy region; because the convergence rates differ, the gap *widens* with budget. The crossing points are summarized in Table III: to reach fidelity 0.99 the best first-order ordering needs 315 rotations, the antithetic schedule 236 and the symmetric schedule 225—a 1.40× reduction. We note that, to reach 0.999, *no* first-order schedule succeeds within the swept budget (entry "—"), while the antithetic and symmetric schedules reach it at 315 and 300 rotations. Most strikingly, the free antithetic schedule, costing exactly $L \cdot r$ rotations, tracks the 2×-per-step symmetric fold on this matched-cost axis to within line width—the second-order accuracy is recovered at first-order cost.

D. The advantage is size independent

A central question for any product-formula improvement is whether it survives at scale. Because the convergence-order argument of Theorem 2 is size independent, it should; we verify this directly. The matrix-

TABLE III. **Rotations to reach a target fidelity (the matched-cost efficiency metric).** For each schedule, the realized rotation count at which the mean fidelity over the 144 instances first reaches 0.9, 0.99, and 0.999; "—" marks a target not reached within the swept step grid. The folded schedules reach 0.99 at a 1.40× smaller budget than the best first-order ordering and are the only schedules to reach 0.999. Transcribed from the `gates_to_target` block of `results/summary.json`.

schedule	$F \geq 0.9$	$F \geq 0.99$	$F \geq 0.999$
random (1st)	157	472	—
coefficient (1st)	118	315	—
clique (1st)	118	315	—
antithetic (2nd, free)	118	236	315
symmetric (2nd)	150	225	300

TABLE IV. **Scaling of the convergence rate and speedup with system size.** For each size n, the fitted convergence rate p of the first-order (1st), antithetic (anti), and symmetric (sym) schedules, the matched-cost rotation speedup to reach $F \geq 0.99$ (best first-order ordering / best fold), and the mean wall-clock time per instance. The first-order rate stays ≈ 2 and the folded rates stay ≈ 4 at every size, while the matrix-free reference keeps the per-instance cost modest up to $n = 14$. Transcribed from the `by_size` block of `results/summary.json`.

n	p (1st)	p (anti)	p (sym)	speedup	time (s)
4	1.98	4.08	4.03	1.68	0.081
6	1.97	4.07	4.02	2.00	0.121
8	1.97	4.06	4.02	1.58	0.198
10	1.94	4.08	4.03	2.09	0.293
12	1.96	4.08	4.03	2.07	0.731
14	1.94	4.06	4.02	2.06	2.258

free Krylov reference removes the dense-exponential bottleneck and lets us evaluate the exact benchmark from $n = 4$ up to $n = 14$ qubits, the same range over which a dense e^{-iHt} would be intractable. As seen in Fig. 5(a), the fitted convergence rate is flat in n: the three first-order schedules hold a rate ≈ 2 at every size (rate 1.94 at $n = 14$), while the antithetic and symmetric folds hold a rate ≈ 4 (antithetic rate 4.06 at $n = 14$). The doubling of the convergence order is therefore not a small-system artifact but a property of the schedule. Figure 5(b) and Table IV show the same conclusion in the matched-cost metric: the free antithetic fold reaches fidelity 0.99 at fewer rotations than the best first-order ordering at *every* size, with the speedup reaching 2.06× at $n = 14$. This scaling is obtained with the antithetic schedule's identical $L \cdot r$ rotation count, so the advantage is recovered at zero overhead independently of system size.

TABLE V. **Per-family rotations to a target fidelity.** For each family and schedule, the realized rotation count to reach mean fidelity ≥ 0.99 and ≥ 0.999; "—" marks a target not reached within the swept step grid. The folded schedules are the only ones to reach 0.999 on the molecular-like and transverse-field Ising families; the clique-based first-order schedules already reach it at minimal cost on Heisenberg. Transcribed from the `gates_to_target` block of `results/summary.json`.

family	schedule	$F \geq 0.99$	$F \geq 0.999$
heisenberg	random (1st)	—	—
	coefficient (1st)	24	24
	clique (1st)	24	24
	antithetic (2nd, free)	24	24
	symmetric (2nd)	47	47
molecular_like	random (1st)	432	—
	coefficient (1st)	432	—
	clique (1st)	432	—
	antithetic (2nd, free)	216	432
	symmetric (2nd)	205	409
tfim	random (1st)	408	—
	coefficient (1st)	272	—
	clique (1st)	272	—
	antithetic (2nd, free)	136	272
	symmetric (2nd)	129	257

E. Instance dependence and the learned scheduler

The gain is not uniform across families, and CGS is designed to exploit exactly this structure. Figure 6 and Table V decompose the rotations-to-target by family. On the dense *molecular-like* and *transverse-field Ising* families the folded schedules cut the cost to reach 0.99 by roughly a third to a half and are the only schedules to reach 0.999. On *Heisenberg*, by contrast, the commuting-clique structure makes the first-order clique schedule essentially *exact* at a single step (Proposition 2): every clique-based schedule reaches 0.999 at the minimal rotation count, and the symmetric fold's 2× per-step overhead is here wasted. The gate-optimal schedule is therefore genuinely instance dependent—fold the hard instances, and leave the fast-forwardable ones first-order.

The learned scheduler of Sec. II F exploits this structure end-to-end. Trained leave-one-instance-out over the ten commutator-graph features, so that no test instance informs its own predictor, it attains 93% accuracy against the per-instance oracle. We note that, on the subset of 92 instances where every fixed schedule reaches the target, the scheduler needs 114.0 rotations on average, against an oracle optimum of 112.4 and the best single fixed schedules' 117.0 (symmetric) and 117.6 (antithetic)—recovering near-oracle efficiency by routing fast-forwardable instances to the cheap first-order schedule and hard instances to a fold. The scheduler is interpretable and listing-order invariant by construction (Proposition 3): it is a selector over named schedules,

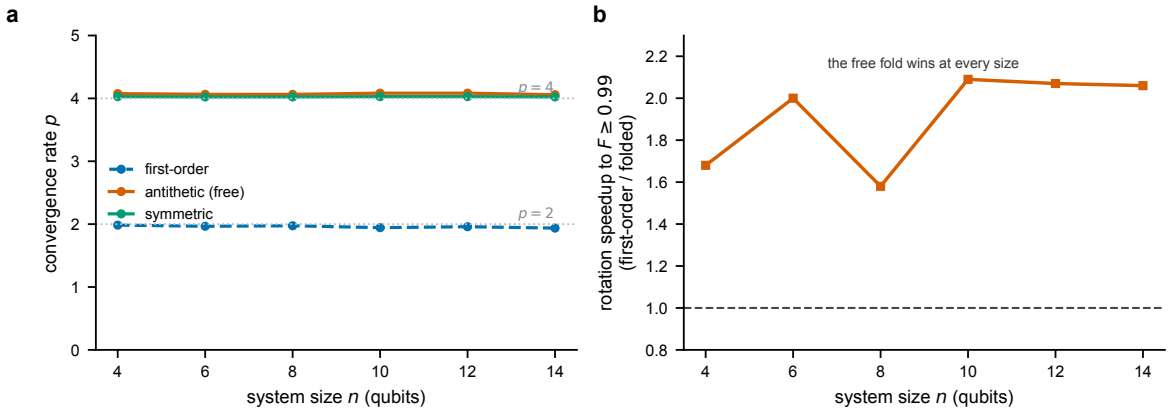

FIG. 5. **The convergence-order doubling and the matched-cost speedup are size independent.** (a) Fitted convergence rate p (mean per-instance log–log slope of infidelity vs. r) versus system size n. The first-order schedule (blue, dashed) holds $p \approx 2$ and the antithetic (vermillion) and symmetric (green) folds hold $p \approx 4$ across the whole range $n = 4$–14; the dotted guides mark $p = 2$ and $p = 4$. (b) Matched-cost rotation speedup to reach $F \geq 0.99$ (best first-order ordering divided by the best fold) versus n; the free fold wins at every size, reaching 2.06× at $n = 14$. The matrix-free Krylov reference, validated against dense `expm` to 10^{-9}, is what makes the exact benchmark tractable to 14 qubits.

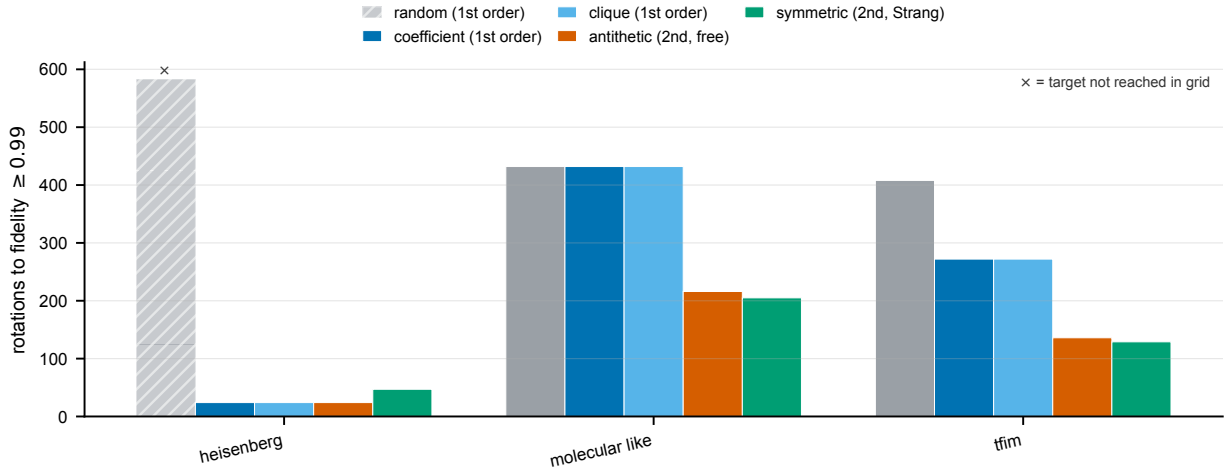

FIG. 6. **Rotations to fidelity ≥ 0.99 by Hamiltonian family.** Lower is better; an × marks a target not reached within the swept budget. On the dense molecular-like and transverse-field Ising families the folded schedules (antithetic, symmetric) reach the target at substantially fewer rotations than any first-order schedule. On Heisenberg the commuting-clique first-order schedules are already near-exact at the minimal budget, so folding only adds overhead. Bars are over the $N = 48$ instances per family.

not a black-box policy.

V. DISCUSSION

a. Summary of the result. In this manuscript we introduced commutator-graph scheduling (CGS), and the result it establishes has two halves that must be kept together. *Negatively*, a single term ordering is a provably weak lever for first-order Trotterization: reordering only re-signs the leading-error commutators, the leading-error norm is exactly invariant on collision-free families and within a few percent of its floor otherwise, and no ordering escapes the $O(t^2/r)$ rate—turning the well-known empirical observation that ordering heuristics are within noise into a theorem. *Positively*, the freedom that does matter is the per-step schedule: the antithetic fold cancels the leading error and converges at second order at the identical gate count, reaching a target fidelity with 1.40× fewer rotations and reaching 0.999 where first-order Trotter plateaus. We do not claim a new product-formula *primitive*—the symmetrization mechanism is classical [4, 7]—but rather the framing that the lever is the schedule, that a fixed ordering provably cannot realize it, and that the commutator graph both certifies the impotence and structures the schedule, together with the matched-cost quantification and the learned scheduler.

b. Relation to prior work. The commutator graph and its clique structure are established in measurement grouping for variational algorithms [12–14]; we repurpose them for *time evolution* and, beyond ordering, for scheduling. The sensitivity of Trotterized chemistry to term ordering has been reported empirically [10, 11, 15], and the impotence theorem explains the ceiling those studies encounter. Symmetric (Strang) and higher-order product formulas are classical [4, 7, 27, 28], and randomized compilation—random permutation per step, qDRIFT, stochastic sparsification, randomized multi-product formulas, and composite channels—improves error scaling by averaging rather than cancellation [8, 9, 43, 44, 46]; the antithetic schedule is the *deterministic, zero-variance* counterpart that cancels the same leading term in a single sample, and our matched-cost frontier places it against both the first-order and the randomized baselines. A complementary deterministic route reduces the leading error by taking linear combinations of Trotter circuits—multi-product formulas [45]—at the cost of ancillas or postselection, whereas the antithetic schedule pays no such overhead. Beyond product formulas, query-optimal primitives based on linear combinations of unitaries, truncated Taylor series, quantum signal processing, and qubitization achieve logarithmic precision scaling [29, 32–35], and refined input- and observable-dependent analyses tighten the cost further [47, 51]; these target the fault-tolerant regime, while CGS is a near-term, zero-overhead upgrade. Hardware-aware compilations such as the fermionic-swap network reduce circuit depth for the same Trotter step [50]. The tight commutator-scaling theory of Trotter error [5, 6] is the lens throughout, and the commutator error form is its explicit, matrix-free, ordering-resolved representation.

c. Why a matched-cost schedule matters. For first-order formulas the schedule is genuinely free—it changes no gate, qubit, or depth count—so any accuracy it recovers is pure profit. In the high-accuracy regime where precise chemistry and materials simulation operate, the difference between an $O(t^2/r)$ and an $O(t^3/r^2)$ method at fixed cost is the difference between reaching and not reaching a target, as the unreachable 0.999 column of Table III shows. The commutator graph supplies both halves of the recommendation at no extra cost: it certifies when ordering cannot help, and its clique coloring tells the learned scheduler which instances to fold and which to leave alone.

d. Limitations. The boundaries of the study are explicit. (i) The second-order advantage is a small-step-size effect: at coarse steps the antithetic fold can be *worse* than a good first-order ordering (its curve in Fig. 4 starts above and crosses below near the intermediate budget); we report the crossover. (ii) The matrix-free Krylov reference extends the exact benchmark to $n = 14$ qubits—and we verify there that the advantage is size independent (Sec. IV D)—but truly large systems ($n \gtrsim 30$), where even a sparse statevector is intractable and a tensor-network reference is required, are not reached; the exact *spectral-norm* impotence sweep is itself capped at $n \leq 8$ (the matrix-free collision and Hilbert–Schmidt statistics carry Theorem 1 at all sizes). (iii) The families are transverse-field Ising, Heisenberg, and random local two-body molecular-like sums, a proxy for, not an instance of, mapped molecular electronic-structure Hamiltonians at scale [17, 18]. (iv) The cost is the realized rotation count and ignores the gate-synthesis cost of individual rotations and hardware connectivity; the antithetic schedule's seam rotations could be merged to reduce the count further, which we do not exploit so that its budget matches first-order exactly. (v) All results are classical statevector simulations on CPU, with no hardware validation.

e. Outlook. Most critically, because the antithetic upgrade changes no gate, qubit, or depth count, commutator-graph scheduling constitutes a zero-overhead accuracy improvement that any first-order Trotter compilation can adopt unchanged on near-term hardware. Having shown that the advantage is size independent up to $n = 14$ qubits with a matrix-free reference, natural next steps are truly large systems with a tensor-network reference, real mapped molecular Hamiltonians, higher-order folded schedules, and a graph neural policy trained directly on a differentiable matched-cost objective. Two of these directions are statistical at their core and worth stating as such. The learned scheduler (Sec. II F) is already a *probabilistic surrogate* for an expensive simulator: rather than run every schedule on every instance, it predicts the gate-optimal choice from cheap, interpretable commutator-graph features whose invariance we prove (Proposition 3), and the same construction invites calibrated uncertainty on that prediction and active-learning selection of which instances to simulate next—an experimental design over training data that our leave-one-out protocol only begins to exploit. In parallel, the shot-based certification of the residual error (Theorem 6) is an estimation problem on hardware, and pushing it to noisy devices—where the witness must be disentangled from device error by the same estimation-plus-hypothesis-testing logic that underlies randomized benchmarking—would let CGS's zero-overhead accuracy claim be validated experimentally rather than by simulation alone. In short, by recasting the leading Trotter error as a certifiable operator on the commutator graph, CGS reframes term ordering as the weak lever it provably is, and identifies the per-step schedule as the free, structured degree of freedom that simulation should exploit.

Appendix A: Proofs for the error-operator theory

We collect the two proofs deferred from Sec. III: the quantitative nested-commutator remainder (Lemma 4) and the full NP-hardness reduction (Theorem 5).

1. Proof of Lemma 4 (quantitative BCH/Magnus remainder)

Write $A(\Delta) = \prod_{a=1}^{m} e^{-\mathrm{i}\Delta B_a}$ with Hermitian B_a, $\|B_a\| = |b_a|$ (a Pauli string has unit operator norm), and $\mu = \sum_a |b_a|$. Introduce the interpolation $A(\Delta, s) = \prod_{a=1}^{m} e^{-\mathrm{i}s\Delta B_a}$, $s \in [0,1]$, so $A(\Delta) = A(\Delta, 1)$, and let $A(\Delta, s) = \exp(\Omega(\Delta, s))$ be the (convergent, for small Δ) Magnus exponent with $\Omega(\Delta, 0) = 0$. Differentiating the ordered product in s and using the Magnus differential equation $\dot{\Omega} = \sum_{k\geq 0} \frac{B_k}{k!} \mathrm{ad}_\Omega^k(\dot{A}A^{-1})$ (with B_k the Bernoulli numbers and $\mathrm{ad}_X(Y) = [X, Y]$), the generator is the norm-convergent series

$$\Omega(\Delta, 1) = \sum_{k\geq 1} \Omega_k \Delta^k,$$

where each Ω_k is a fixed linear combination of k-fold nested commutators $[B_{a_1}, [B_{a_2}, [\ldots, B_{a_k}]\ldots]]$. Two elementary bounds control the tail. First, each nested commutator of order k satisfies, by the triangle inequality $\|[X, Y]\| \leq 2\|X\|\,\|Y\|$ applied $k-1$ times and submultiplicativity across the ordered product,

$$\Big\| k\text{-fold commutator of the } B_a \Big\| \;\leq\; (2)^{k-1}\Big(\sum_a |b_a|\Big)^k = 2^{k-1}\mu^k.$$

Second, the absolute sum of the scalar Magnus/BCH coefficients at order k is dominated by the coefficients of the *scalar* majorant series obtained by replacing every $\|[X, Y]\|$ by $2\|X\|\,\|Y\|$; that majorant is exactly the series for $-\log(2 - e^{2\mu|\Delta|})$ about $\Delta = 0$, which converges for $2\mu|\Delta| < \log 2$. Truncating at order p and setting $G_p(\Delta) = \mathrm{i}\sum_{k\leq p} \Omega_k \Delta^k$ (Hermitian since each Ω_k is anti-Hermitian, being a real combination of commutators of Hermitian operators times $-\mathrm{i}$), the remainder $\Delta^{p+1}\mathcal{R}_p(\Delta) = \sum_{k>p} \Omega_k \Delta^k$ is bounded by the tail of the majorant:

$$\|\mathcal{R}_p(\Delta)\| \leq |\Delta|^{-(p+1)} \sum_{k>p} \frac{(2\mu|\Delta|)^k}{k} \;\leq\; \frac{(2\mu)^{p+1}}{p+1} \sum_{j\geq 0} (2\mu|\Delta|)^j = \frac{(2\mu)^{p+1}}{p+1} \cdot \frac{1}{1-2\mu|\Delta|}$$

and $1/(1-2\mu|\Delta|) \leq e^{2\mu|\Delta|}$ for $2\mu|\Delta| < \log 2 < 1$, giving Eq. (8). The order-1 term is $\Omega_1\Delta = -\mathrm{i}\Delta\sum_a B_a$ (each factor contributes its generator linearly), so G_p's leading part is $\Delta\sum_a B_a$; the classical order-2 BCH term of an ordered product is $-\frac{\Delta^2}{2}\sum_{a<b}[B_a, B_b]$ in the exponent, i.e. the stated $-\frac{\Delta^2}{2}\sum_{a<b}\mathrm{i}[B_a, B_b]$ inside $-\mathrm{i}G_p$. □

2. Full reduction for Theorem 5

Let $G = (V, E_G)$ be an instance of Max-Cut. We construct, in time polynomial in $|V| + |E_G|$, a term list $\{(c_j, P_j)\}$ whose MinResidual objective F equals $2|E_G| - 2\sum_{uv\in E_G} x_u x_v$ under a bijection between sign assignments and cuts, so that minimizing F solves the (NP-hard) minimum-agreement problem, equivalently maximizing the cut.

Gadget. Assign to each vertex $u \in V$ a Boolean sign $x_u \in \{\pm 1\}$, encoded by the relative order of a designated "anchor" term A and a per-vertex term T_u in the schedule: $x_u = \varepsilon_{A,u} \in \{\pm 1\}$ is exactly the sign the ordering places on the anchor–T_u edge (any acyclic tournament realizes an arbitrary such sign pattern, so all $x \in \{\pm 1\}^V$ are feasible). For each graph edge $uv \in E_G$ introduce two CGS-edges that land on a *shared* Pauli string R_{uv} (achievable by choosing T_u, T_v, A so the two anticommuting products $P_A P_{T_u}$ and $P_A P_{T_v}$ coincide up to phase—a collision, in the sense of Definition 1), with unit weights $w = 1$ and base signs s chosen so that the group amplitude is $\alpha_{R_{uv}} = x_u - x_v$. Distinct graph edges use disjoint collision groups, so

$$F(\varepsilon) = \sum_{uv\in E_G} \alpha_{R_{uv}}^2 = \sum_{uv\in E_G} (x_u - x_v)^2.$$

Correctness. Using $x_u^2 = 1$, $(x_u - x_v)^2 = 2 - 2x_u x_v \in \{0, 4\}$: it is 0 when $x_u = x_v$ (endpoints uncut) and 4 when $x_u \neq x_v$ (endpoints cut). Hence $F = 4 \cdot |\{uv : \text{cut}\}|$ and $\frac{1}{4}F = |E_G| - \frac{1}{2}\sum_{uv} x_u x_v$. Therefore a sign assignment minimizing F minimizes the number of cut edges, equivalently *maximizes* the agreement $\sum_{uv} x_u x_v$; the complementary maximization of F is Max-Cut. Because computing Max-Cut (and its Min-Uncut complement) is NP-hard and the reduction is polynomial, exactly minimizing F—i.e. solving MinResidual—is NP-hard. The construction never forms a $2^n \times 2^n$ matrix: each collision is verified by the $O(n)$ Pauli-product kernel of Sec. II A, so the reduction is explicit and efficient. □

Appendix B: Numerical methods and protocol

The analytic development—the Pauli algebra, the commutator error form, and the ordering-impotence, antithetic-cancellation, clique-exactness, and listing-order-invariance results—appears with complete proofs in Sec. II. Here we specify the numerical protocol. Three families are generated as $\{(c_j, P_j)\}$ term lists: transverse-field Ising $H = -J\sum_i Z_i Z_{i+1} - h\sum_i X_i$; Heisenberg $H = \sum_i (J_x X_i X_{i+1} + J_y Y_i Y_{i+1} + J_z Z_i Z_{i+1})$; and *molecular-like*, random local two-body Pauli terms with seeded Gaussian coefficients—a tractable stand-in for the dense, irregular anticommutation structure of Jordan–Wigner / Bravyi–Kitaev mapped electronic-structure Hamiltonians [17, 18, 49, 52–54]. The reported-scale configuration (`configs/full.yaml`) uses sizes $n \in \{4, 6, 8, 10, 12, 14\}$, eight seeded instances per (family, size), evolution time $t = 1.0$, a reference budget of $r = 6$ steps, and a step grid $\{1, 2, 4, 6, 8, 12, 16, 24\}$ for the convergence and frontier curves, giving 144 instances in total. The exact reference $e^{-\mathrm{i}Ht}|\psi_0\rangle$ is computed either by dense `scipy.linalg.expm` (the ground truth, used for $n \leq 8$) or by the *matrix-free* Krylov reference `expm_multiply` applied to a sparse Hamiltonian assembled term-by-term from one-nonzero-per-column Pauli permutations [Eq. (4) extended to the reference]; the latter never forms a

$2^n \times 2^n$ matrix, agrees with dense `expm` to 10^{-9} (the `reference_backend_verified` flag), and is what extends the exact benchmark to $n = 14$. For each instance and schedule we evaluate the product formula, score fidelity and infidelity against the reference, and record the realized rotation count; from the (rotation count, fidelity) curves we read the rotations to each target fidelity and fit the convergence rate. Ordering impotence is measured by sampling 48 orderings per instance and forming the exact leading-error spectral norm on the 72 instances with $n \leq 8$, where the dense $O((2^n)^3)$ spectral norm is tractable; the matrix-free Hilbert–Schmidt surrogate and collision statistics are computed at all sizes. Per-schedule means carry 95% confidence intervals $1.96\,\hat{\sigma}/\sqrt{N}$.

Appendix C: Software, validation, and reproducibility

The pipeline is a CPU-only Python package (`topoham`) depending on NumPy [19], SciPy [20], NetworkX [21], scikit-learn [16], and Matplotlib. Hamiltonians are stored as Pauli term lists (`hamiltonians.py`); the Pauli algebra of Sec. II A—the $O(n)$ parity test (Lemma 1), the cosine–sine rotation kernel (Lemma 2), and the phased-Pauli product—is the inner loop of the exact statevector simulator (`pauli.py`). The commutator graph and matrix-free error form are built in `commutator_graph.py` and `error_form.py`: the greedy coloring partitions terms into commuting cliques (Proposition 2), the error form assembles Eq. (5) as a signed Pauli-string sum over the edges (Definition 1), and the same module computes the ten listing-order-invariant features the learned scheduler reads (Proposition 3). Schedules are realized as flat lists of (term, step-fraction) rotations whose seam-merged length is the gate cost on which all schedules are compared (`schedules.py`, `env.py`); the learned scheduler (`LearnedSchedulePolicy`) is the standardized logistic-regression classifier of Sec. II F, trained leave-one-instance-out. The exact reference is built in `env.py`: `sparse_pauli_matrix` assembles each Pauli string as a one-nonzero-per-column signed permutation directly from index arithmetic in $O(2^n)$, so the sparse Hamiltonian and its `expm_multiply` reference are formed without a dense $2^n \times 2^n$ matrix.

The single source of truth is `results/summary.json`, written by the runner with full provenance (seed, command, platform, package versions, runtime, peak memory) and *three* integrity flags that must pass before any fidelity is produced: `pauli_algebra_verified` (the fast kernel matches dense `expm` to 10^{-9}), `error_form_verified` (the matrix-free error form matches the dense pair-commutator sum to 10^{-9}), and `reference_backend_verified` (the matrix-free Krylov reference matches dense `expm` to 10^{-9} before it is used at scale). On the run reported here (seed 0, Python 3.13, macOS arm64), the experiment completed in ≈ 209 s wall-clock with ≈ 500 MB peak memory, all three flags `true`. The macros, tables, and figures in this manuscript are generated directly from `summary.json`, so no number is entered by hand. The full pipeline is reproduced from the accompanying code by

```
pip install .
make test
export KMP_DUPLICATE_LIB_OK=TRUE
topoham-reproduce \
  --config configs/full.yaml
```

where `make test` checks the Pauli kernel (10^{-9}), the error form (10^{-9}), antithetic second-order convergence, and clique exactness, and `topoham-reproduce` regenerates `summary.json`, the tables, and the figures. All experiments are reproducible on commodity hardware.

[1] Richard P. Feynman. Simulating physics with computers. *International Journal of Theoretical Physics*, 21:467–488, 1982. doi:10.1007/BF02650179.

[2] Seth Lloyd. Universal quantum simulators. *Science*, 273(5278):1073–1078, 1996. doi:10.1126/science.273.5278.1073.

[3] Hale F. Trotter. On the product of semi-groups of operators. *Proceedings of the American Mathematical Society*, 10(4):545–551, 1959. doi:10.1090/S0002-9939-1959-0108732-6.

[4] Masuo Suzuki. General theory of fractal path integrals with applications to many-body theories and statistical physics. *Journal of Mathematical Physics*, 32(2):400–407, 1991. doi:10.1063/1.529425.

[5] Andrew M. Childs, Yuan Su, Minh C. Tran, Nathan Wiebe, and Shuchen Zhu. Theory of trotter error with commutator scaling. *Physical Review X*, 11(1):011020, 2021. doi:10.1103/PhysRevX.11.011020.

[6] Andrew M. Childs, Dmitri Maslov, Yunseong Nam, Neil J. Ross, and Yuan Su. Toward the first quantum simulation with quantum speedup. *Proceedings of the National Academy of Sciences*, 115(38):9456–9461, 2018. doi:10.1073/pnas.1801723115.

[7] Naomichi Hatano and Masuo Suzuki. Finding exponential product formulas of higher orders. *Lecture Notes in Physics*, 679:37–68, 2005. doi:10.1007/11526216_2.

[8] Andrew M. Childs, Aaron Ostrander, and Yuan Su. Faster quantum simulation by randomization. *Quantum*, 3:182, 2019. doi:10.22331/q-2019-09-02-182.

[9] Earl Campbell. Random compiler for fast Hamiltonian simulation. *Physical Review Letters*, 123(7):070503, 2019. doi:10.1103/PhysRevLett.123.070503.

[10] David Poulin, Matthew B. Hastings, Dave Wecker, Nathan Wiebe, Andrew C. Doherty, and Matthias Troyer. The trotter step size required for accurate quantum simulation of quantum chemistry. *Quantum Information & Computation*, 15(5–6):361–384, 2015.

[11] Matthew B. Hastings, Dave Wecker, Bela Bauer, and

Matthias Troyer. Improving quantum algorithms for quantum chemistry. *Quantum Information & Computation*, 15(1–2):1–21, 2015.

[12] Vladyslav Verteletskyi, Tzu-Ching Yen, and Artur F. Izmaylov. Measurement optimization in the variational quantum eigensolver using a minimum clique cover. *The Journal of Chemical Physics*, 152(12):124114, 2020. doi: 10.1063/1.5141458.

[13] Pranav Gokhale, Olivia Angiuli, Yongshan Ding, Kaiwen Gui, Teague Tomesh, Martin Suchara, Margaret Martonosi, and Frederic T. Chong. $O(N^3)$ measurement cost for variational quantum eigensolver on molecular hamiltonians. *IEEE Transactions on Quantum Engineering*, 1:1–24, 2020. doi:10.1109/TQE.2020.3035814.

[14] Ophelia Crawford, Billy van Straaten, Daochen Wang, Thomas Parks, Earl Campbell, and Stephen Brierley. Efficient quantum measurement of pauli operators in the presence of finite sampling error. *Quantum*, 5:385, 2021. doi:10.22331/q-2021-01-20-385.

[15] Andrew Tranter, Peter J. Love, Florian Mintert, and Nathan Wiebe. Ordering of trotterization: Impact on errors in quantum simulation of electronic structure. *Entropy*, 21(12):1218, 2019. doi:10.3390/e21121218.

[16] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, et al. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[17] Sergey B. Bravyi and Alexei Yu. Kitaev. Fermionic quantum computation. *Annals of Physics*, 298(1):210–226, 2002. doi:10.1006/aphy.2002.6254.

[18] James D. Whitfield, Jacob Biamonte, and Alán Aspuru-Guzik. Simulation of electronic structure hamiltonians using quantum computers. *Molecular Physics*, 109(5): 735–750, 2011. doi:10.1080/00268976.2011.552441.

[19] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, et al. Array programming with numpy. *Nature*, 585:357–362, 2020. doi:10.1038/s41586-020-2649-2.

[20] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, et al. Scipy 1.0: fundamental algorithms for scientific computing in python. *Nature Methods*, 17:261–272, 2020. doi: 10.1038/s41592-019-0686-2.

[21] Aric A. Hagberg, Daniel A. Schult, and Pieter J. Swart. Exploring network structure, dynamics, and function using networkx. In *Proceedings of the 7th Python in Science Conference*, pages 11–15, 2008.

[22] I. M. Georgescu, S. Ashhab, and Franco Nori. Quantum simulation. *Reviews of Modern Physics*, 86(1):153–185, 2014. doi:10.1103/RevModPhys.86.153.

[23] Alán Aspuru-Guzik, Anthony D. Dutoi, Peter J. Love, and Martin Head-Gordon. Simulated quantum computation of molecular energies. *Science*, 309(5741):1704–1707, 2005. doi:10.1126/science.1113479.

[24] Sam McArdle, Suguru Endo, Alán Aspuru-Guzik, Simon C. Benjamin, and Xiao Yuan. Quantum computational chemistry. *Reviews of Modern Physics*, 92(1): 015003, 2020. doi:10.1103/RevModPhys.92.015003.

[25] B. P. Lanyon, C. Hempel, D. Nigg, M. Müller, R. Gerritsma, F. Zähringer, P. Schindler, J. T. Barreiro, M. Rambach, G. Kirchmair, M. Hennrich, P. Zoller, R. Blatt, and C. F. Roos. Universal digital quantum simulation with trapped ions. *Science*, 334(6052):57–61, 2011. doi:10.1126/science.1208001.

[26] J. Smith, A. Lee, P. Richerme, B. Neyenhuis, P. W. Hess, P. Hauke, M. Heyl, D. A. Huse, and C. Monroe. Many-body localization in a quantum simulator with programmable random disorder. *Nature Physics*, 12(10):907–911, 2016. doi:10.1038/nphys3783.

[27] Gilbert Strang. On the construction and comparison of difference schemes. *SIAM Journal on Numerical Analysis*, 5(3):506–517, 1968. doi:10.1137/0705041.

[28] Masuo Suzuki. Fractal decomposition of exponential operators with applications to many-body theories and Monte Carlo simulations. *Physics Letters A*, 146(6):319–323, 1990. doi:10.1016/0375-9601(90)90962-N.

[29] Dominic W. Berry, Graeme Ahokas, Richard Cleve, and Barry C. Sanders. Efficient quantum algorithms for simulating sparse Hamiltonians. *Communications in Mathematical Physics*, 270(2):359–371, 2007. doi: 10.1007/s00220-006-0150-x.

[30] Andrew M. Childs and Nathan Wiebe. Hamiltonian simulation using linear combinations of unitary operations. *Quantum Information & Computation*, 12(11–12):901–924, 2012.

[31] Andrew M. Childs and Robin Kothari. Limitations on the simulation of non-sparse Hamiltonians. *Quantum Information & Computation*, 10(7–8):669–684, 2010.

[32] Dominic W. Berry, Andrew M. Childs, Richard Cleve, Robin Kothari, and Rolando D. Somma. Simulating Hamiltonian dynamics with a truncated Taylor series. *Physical Review Letters*, 114(9):090502, 2015. doi: 10.1103/PhysRevLett.114.090502.

[33] Dominic W. Berry, Andrew M. Childs, and Robin Kothari. Hamiltonian simulation with nearly optimal dependence on all parameters. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 792–809. IEEE, 2015. doi: 10.1109/FOCS.2015.54.

[34] Guang Hao Low and Isaac L. Chuang. Optimal Hamiltonian simulation by quantum signal processing. *Physical Review Letters*, 118(1):010501, 2017. doi: 10.1103/PhysRevLett.118.010501.

[35] Guang Hao Low and Isaac L. Chuang. Hamiltonian simulation by qubitization. *Quantum*, 3:163, 2019. doi: 10.22331/q-2019-07-12-163.

[36] Andrew M. Childs and Yuan Su. Nearly optimal lattice simulation by product formulas. *Physical Review Letters*, 123(5):050503, 2019. doi: 10.1103/PhysRevLett.123.050503.

[37] Minh C. Tran, Su-Kuan Chu, Yuan Su, Andrew M. Childs, and Alexey V. Gorshkov. Destructive error interference in product-formula lattice simulation. *Physical Review Letters*, 124(22):220502, 2020. doi: 10.1103/PhysRevLett.124.220502.

[38] David Layden. First-order Trotter error from a second-order perspective. *Physical Review Letters*, 128(21): 210501, 2022. doi:10.1103/PhysRevLett.128.210501.

[39] Changhao Yi and Elizabeth Crosson. Spectral analysis of product formulas for quantum simulation. *npj Quantum Information*, 8(1):37, 2022. doi:10.1038/s41534-022-00548-w.

[40] Chi-Fang Chen and Fernando G. S. L. Brandão. Average-case speedup for product formulas. *Communications in Mathematical Physics*, 405(2):32, 2024. doi: 10.1007/s00220-023-04912-5.

[41] Markus Heyl, Philipp Hauke, and Peter Zoller. Quantum localization bounds Trotter errors in digital quantum simulation. *Science Advances*, 5(4):eaau8342, 2019. doi:10.1126/sciadv.aau8342.

[42] L. M. Sieberer, T. Olsacher, A. Elben, M. Heyl, P. Hauke,

F. Haake, and P. Zoller. Digital quantum simulation, Trotter errors, and quantum chaos of the kicked top. *npj Quantum Information*, 5(1):78, 2019. doi:10.1038/s41534-019-0192-5.

[43] Yingkai Ouyang, David R. White, and Earl T. Campbell. Compilation by stochastic Hamiltonian sparsification. *Quantum*, 4:235, 2020. doi:10.22331/q-2020-02-27-235.

[44] Paul K. Faehrmann, Mark Steudtner, Richard Kueng, Maria Kieferová, and Jens Eisert. Randomizing multi-product formulas for Hamiltonian simulation. *Quantum*, 6:806, 2022. doi:10.22331/q-2022-09-19-806.

[45] Sergiy Zhuk, Niall Robertson, and Sergey Bravyi. Trotter error bounds and dynamic multi-product formulas for Hamiltonian simulation. *Physical Review Research*, 6(3): 033309, 2024. doi:10.1103/PhysRevResearch.6.033309.

[46] Matthew Hagan and Nathan Wiebe. Composite quantum simulations. *Quantum*, 7:1181, 2023. doi:10.22331/q-2023-11-14-1181.

[47] Dong An, Di Fang, and Lin Lin. Time-dependent unbounded Hamiltonian simulation with vector norm scaling. *Quantum*, 5:459, 2021. doi:10.22331/q-2021-05-26-459.

[48] Dave Wecker, Matthew B. Hastings, Nathan Wiebe, Bryan K. Clark, Chetan Nayak, and Matthias Troyer. Solving strongly correlated electron models on a quantum computer. *Physical Review A*, 92(6):062318, 2015. doi:10.1103/PhysRevA.92.062318.

[49] Ryan Babbush, Dominic W. Berry, Ian D. Kivlichan, Annie Y. Wei, Peter J. Love, and Alán Aspuru-Guzik. Exponentially more precise quantum simulation of fermions in second quantization. *New Journal of Physics*, 18(3): 033032, 2016. doi:10.1088/1367-2630/18/3/033032.

[50] Ian D. Kivlichan, Jarrod McClean, Nathan Wiebe, Craig Gidney, Alán Aspuru-Guzik, Garnet Kin-Lic Chan, and Ryan Babbush. Quantum simulation of electronic structure with linear depth and connectivity. *Physical Review Letters*, 120(11):110501, 2018. doi:10.1103/PhysRevLett.120.110501.

[51] Ian D. Kivlichan, Nathan Wiebe, Ryan Babbush, and Alán Aspuru-Guzik. Bounding the costs of quantum simulation of many-body physics in real space. *Journal of Physics A: Mathematical and Theoretical*, 50(30):305301, 2017. doi:10.1088/1751-8121/aa77b8.

[52] Pascual Jordan and Eugene Wigner. Über das Paulische Äquivalenzverbot. *Zeitschrift für Physik*, 47(9–10):631–651, 1928. doi:10.1007/BF01331938.

[53] Jacob T. Seeley, Martin J. Richard, and Peter J. Love. The Bravyi-Kitaev transformation for quantum computation of electronic structure. *The Journal of Chemical Physics*, 137(22):224109, 2012. doi:10.1063/1.4768229.

[54] Andrew Tranter, Sarah Sofia, Jake Seeley, Michael Kaicher, Jarrod McClean, Ryan Babbush, Peter V. Coveney, Florian Mintert, Frank Wilhelm, and Peter J. Love. The Bravyi-Kitaev transformation: Properties and applications. *International Journal of Quantum Chemistry*, 115(19):1431–1441, 2015. doi:10.1002/qua.24969.

[55] Tzu-Ching Yen, Vladyslav Verteletskyi, and Artur F. Izmaylov. Measuring all compatible operators in one series of single-qubit measurements using unitary transformations. *Journal of Chemical Theory and Computation*, 16 (4):2400–2409, 2020. doi:10.1021/acs.jctc.0c00008.

[56] Andrew Jena, Scott Genin, and Michele Mosca. Pauli partitioning with respect to gate sets. *arXiv preprint arXiv:1907.07859*, 2019.

[57] M. Huynh. Query-Efficient Quantum Approximate Optimization via Graph-Conditioned Trust Regions. arXiv:2604.24803 (2026).

[58] M. Huynh. Graph-conditioned trust regions for uncertainty-calibrated quantum approximate optimization (extended). Manuscript in preparation (2026).

[59] M. Huynh. The measurement cost of warm-started low-depth QAOA. Manuscript in preparation (2026).

[60] M. Huynh. Certified query budgets for the quantum approximate optimization algorithm. Manuscript in preparation (2026).

[61] M. Huynh. Topology-Conditioned QAOA Parameter Transfer for Budgeted Graph Optimization (companion). Manuscript in preparation (2026).

[62] M. Huynh. Operator-Spectral Truncated Priors for Query-Efficient QAOA Parameter Search. Manuscript in preparation (2026).

[63] M. Huynh. Spectral-truncation graph kernels for QAOA warm starts: topology-conditioned schedule transfer beyond depth one. Manuscript in preparation (2026).