

QN vs ARS vs ARC

Three frameworks. One paper. One backend. Simple and high-precision Hamiltonian simulation by compensating Trotter error with LCU

Same paper. Same backend. Three pipelines.

June 2026

Project	QuantumNovelty (QN) — audit-and-falsify framework for quantum-computing research
Repository	https://github.com/boltzmannentropy/QuantumNovelty
Author	Shlomo Kashani (QNeura.ai)
Paper under review	arXiv:2212.04566 — prx-quantum
LLM backend	Claude Code CLI (2.1.1 (Claude Code))
Model snapshots used	claude-haiku-4-5-20251001 claude-opus-4-5-20251101
Report generated	2026-06-10 20:45 by <code>build_compare_report.py</code>
Run directory	<code>examples/end_to_end/compare_qn_vs_ars_vs_arc/_run/</code>

Setup

The paper — *Simple and high-precision Hamiltonian simulation by compensating Trotter error with LCU* — was run through three review frameworks on the same backend and the same extracted text, so output differences reflect each framework’s design rather than LLM variance. The per-stage model snapshot, exact token counts, and USD cost (from the Claude CLI’s JSON envelope) appear in the ledger below.

QN pipeline: `chain/run.sh -pipeline paper-audit` — 4 stages (research / reviewer / fallacies / cqe).

ARS pipeline: `academic-paper-reviewer` (full mode) — 7-agent orchestration (Phase 0 field-analyst, Phase 1a–e five reviewers, Phase 2 editorial synthesizer).

ARC pipeline: `peer_review + quality_gate` stages — 2-stage review subset extracted from ARC’s 23-stage full `python -m researchclaw run pipeline`.

1 Pipeline architecture map

How the two frameworks line up. Both run on the same paper text and the same backend; differences in output reflect each framework’s design choices.

QN paper-audit	ARS academic-paper-reviewer	ARC peer_review + quality_gate
<p>Stage 1: deep_research -mode review audit-and-falsify checklist scored against paper text (15 items: novelty rigour, baseline coverage, ablations, Wilson CIs, etc.) <i>Output:</i> research_quality_review.md</p>	<p>Phase 0: field_analyst_agent reads the paper, identifies field, configures 5 reviewer personas with named expertise. <i>Output:</i> reviewer configuration cards (x5)</p>	<p>(no dedicated equivalent — topic context is injected via prompt placeholder in both ARC stages.)</p>
<p>Stage 2 (multi-voice): quantum_reviewer -mode full single LLM call yielding 5-voice panel: EIC + R1 (Physics) + R2 (Novelty) + R3 (Evidence) + Devil’s Advocate + reconciliation. <i>Output:</i> review_panel.md (one file)</p>	<p>Phase 1 (5 separate calls): eic_agent methodology_reviewer_agent domain_reviewer_agent perspective_reviewer_agent devils_advocate_reviewer_agent <i>Output:</i> 5 separate review cards (one per agent)</p>	<p>Stage 1: peer_review single LLM call yielding 2 reviewer perspectives (Reviewer A, Reviewer B) with strengths / weaknesses / actionable revisions + methodology-evidence consistency check + fabrication-flag scan. <i>Output:</i> 01_peer_review.md</p>
<p>Stage 3: logical_fallacies 11 quantum-CS-specific fallacies (cherry-picked-baseline, ad-hoc-precision-floor, simulator-laundering, pareto-cherry-picked-axes, cross-llm-theatre, ...) + standard taxonomy. <i>Output:</i> fallacy_report.md + fallacy_findings.json</p>	<p>(no dedicated equivalent — fallacy work is folded into the Devil’s Advocate agent in Phase 1)</p>	<p>(no dedicated equivalent — fabrication detection is folded into the peer_review prompt’s methodology-evidence consistency check)</p>
<p>Stage 4: process_summary (Stage-6 CQE) 6-dim Collaboration Quality Evaluation, geometric mean composite. Mechanical rubric scoring. <i>Output:</i> cqe_scores.json + process_summary.md</p>	<p>Phase 2: editorial_synthesizer_agent aggregates all 5 review cards, identifies consensus + disagreements, makes editorial decision. <i>Output:</i> editorial decision letter + revision roadmap</p>	<p>Stage 2: quality_gate single LLM call producing JSON verdict: {score_1_to_10, verdict, strengths, weaknesses, required_actions}. Numeric threshold gate (default 7.0) decides PROCEED vs REVISE. <i>Output:</i> 02_quality_gate.json</p>

2 Token + cost ledger — all three frameworks, same backend

Framework	Stage	Model	Input tk	Output tk	Cost (\$)	Elapsed (s)
QN	Deep-research review	claude-opus-4-5-20251101	1	1,420	\$0.5585	38
QN	Reviewer panel 5 voices	claude-haiku-4-5-20251001	1	5,729	\$0.7880	132
QN	Logical-fallacy report	claude-opus-4-5-20251101	1	2,644	\$0.7207	50
QN	Stage-6 CQE narrative	claude-opus-4-5-20251101	2	1,207	\$0.0801	33
QN total			5	11,000	\$2.1473	255
ARS	Phase 0: Field analyst	claude-haiku-4-5-20251001	1	2,141	\$0.4092	53
ARS	Phase 1a: EIC	claude-haiku-4-5-20251001	1	2,601	\$0.4248	62
ARS	Phase 1b: Methodology reviewer	claude-opus-4-5-20251101	1	1,562	\$0.4034	39
ARS	Phase 1c: Domain reviewer	claude-haiku-4-5-20251001	1	2,759	\$0.4343	67
ARS	Phase 1d: Perspective reviewer	claude-opus-4-5-20251101	1	2,644	\$0.4328	66
ARS	Phase 1e: Devil's Advocate	claude-haiku-4-5-20251001	1	2,076	\$0.4316	53
ARS	Phase 2: Editorial synthesizer	claude-haiku-4-5-20251001	1	3,754	\$0.5648	91
ARS total			7	17,537	\$3.1010	435
ARC	Stage 1: Peer review	claude-haiku-4-5-20251001	2	2,755	\$0.3529	62
ARC	Stage 2: Quality gate	claude-haiku-4-5-20251001	2	980	\$0.3067	27
ARC total			4	3,735	\$0.6597	89

All rows carry exact token counts and per-call USD from the Claude Code CLI's JSON envelope — nothing is estimated.

3 QN paper-audit — full per-stage prose

Every subsection below is QN output (`_run/qn/`); the per-stage model snapshot is shown under each heading.

3.1 QN — Deep-research review

Model: `claude-opus-4-5-20251101` · Audit-and-falsify checklist QN-specific.

I'll analyze this paper systematically according to the research-quality assessment framework.

3.1.1 1. One-paragraph Summary of Claims

This paper proposes composite Hamiltonian simulation algorithms that combine Trotter formulas with linear-combination-of-unitaries (LCU) methods to achieve both easy implementation and high precision. The authors present two main algorithms: (1) Paired Taylor-Series Compensation (PTSC), which claims to exponentially improve the accuracy scaling of K th-order Trotter formulas by achieving gate complexity $O(t^{1+1/(2K+1)})$ with logarithmic accuracy dependence; and (2) Nested-Commutator Compensation (NCC), which for lattice Hamiltonians claims to achieve gate complexity $O(n^{1+2/(2K+1)} t^{1+1/(2K+1)} t^{-1/(2K+1)})$, providing 3-4 orders of magnitude higher accuracy than optimal Trotter algorithms at the same gate cost. The key innovation is using LCU formulas to compensate Trotter errors while preserving the advantageous commutator scaling of Trotter methods through an “order-pairing” technique based on Euler’s formula.

3.1.2 2. Audit-and-Falsify Checklist

Item	Status	Evidence
Augmented baseline catalog	PASS	Paper compares against 4th-order Trotter with random permutation [12], post-Trotter methods (Taylor series [25], Jacobi-Anger [28], QSP [27]), and nested-commutator bounds from Ref. [20], which represent current state-of-the-art analytical bounds.
Strict-domination comparator	PARTIAL	Fig. 8 shows gate count comparisons at specific values (10§, 10, 10), but tolerances for the comparisons are not calibrated with explicit <code>_abs/_rel</code> bounds; asymptotic scaling claims (O-notation) dominate over precise threshold analysis.
Recompute-from-raw	FAIL	The paper presents gate count estimates in Fig. 8 but does not provide tabulated raw numerical values from which the plotted ratios (“2 orders of magnitude smaller,” “3-4 orders of magnitude higher accuracy”) can be independently verified.
Wilson 95% CIs	NOT-APPLICABLE	The paper presents analytical bounds and asymptotic complexity results rather than empirical small-sample rates; no statistical claims requiring binomial confidence intervals are made.
Cross-LLM falsifiability	NOT-APPLICABLE	No LLM-in-the-loop methods were used; this is a theoretical quantum algorithms paper with analytical proofs.
Honest negatives	PARTIAL	Paper acknowledges that PTSC has worse system-size scaling than Trotter for lattice Hamiltonians (Table I shows $O(n^{2+1/(2K+1)})$ vs $O(n^{1+1/K})$), but lacks a dedicated “Failure Modes” section discussing regimes where the methods underperform or fail.
Simulator precision floor	NOT-APPLICABLE	Paper presents analytical complexity bounds rather than numerical simulation results; no energy comparisons requiring float64 vs complex64 verification are claimed.

Item	Status	Evidence
Auditable claims	FAIL	No reproducible code, scripts, or JSON artifacts are provided; Theorem 1 and Theorem 2’s gate complexity claims cannot be independently verified from raw computational outputs.

3.1.3 3. Overall Assessment

This paper presents mathematically rigorous theoretical work with clearly stated theorems and proofs. The methodology—combining Trotter formulas with LCU compensation via Euler’s formula pairing—is novel and the asymptotic complexity improvements are analytically derived rather than empirically estimated, which strengthens internal validity. However, the paper falls short of contemporary reproducibility standards: the numerical comparisons in Fig. 8 lack underlying raw data tables, the “2-4 orders of magnitude” improvement claims cannot be independently recomputed from provided artifacts, and the analytical bounds from Refs. [12] and [20] used as baselines are taken at face value without explicit numerical verification. The paper would benefit from a supplementary repository containing gate-counting code and explicit numerical examples. For a theory paper, the core mathematical claims appear sound, but the quantitative comparison statements require stronger evidential grounding.

Research Rigour Score: 6/10

The theoretical framework is solid with proper lemmas, propositions, and theorems, but the quantitative claims connecting theory to practice (gate count estimates, order-of-magnitude improvements) lack the audit trail needed for independent verification.

3.1.4 4. Three Highest-Leverage Improvements

- 1. Add a supplementary code repository with gate-counting scripts:** Provide Python/Julia code that computes the gate counts shown in Fig. 8 from the analytical formulas in Theorems 1-2 and the baseline bounds from Refs. [12, 20]. Include explicit parameter values (n , t , ϵ) and output raw numerical tables that can be independently verified, with checksums or hash verification.
- 2. Create a structured claims table with recomputable evidence:** For each quantitative claim (“2 orders of magnitude smaller,” “3-4 orders of magnitude higher accuracy”), provide a table showing: (a) the exact parameter regime, (b) the baseline gate count value, (c) the proposed method’s gate count, (d) the ratio, and (e) the analytical formula used to compute each. This transforms prose claims into auditable numerical statements.
- 3. Add a Limitations/Failure Modes section:** Explicitly characterize regimes where the methods underperform, such as: (a) the crossover point where 0th-order PTSC outperforms higher-order variants for small t , (b) the overhead coefficient in the t sampling cost (currently buried in Proposition 1), and (c) conditions under which the nested-commutator padding introduces significant constant-factor overhead. This demonstrates honest engagement with method boundaries.

3.2 QN — Reviewer panel 5 voices

Model: claude-haiku-4-5-20251001 · EIC + R1 (Physics) + R2 (Novelty) + R3 (Evidence) + Devil’s Advocate, rolled into one stage.

3.2.1 Voice 1 — Reviewer 1 (Physics correctness)

This manuscript presents a compelling hybrid approach to Hamiltonian simulation that combines Trotter formulas with linear combination of unitaries (LCU) compensation. The physical setup is well-grounded in the standard formalism of quantum simulation, where the Hamiltonian is decomposed as $H = \sum H_i$ with each summand being a weighted Pauli operator. The authors correctly identify the fundamental tension between Trotter methods, which enjoy commutator scaling and near-linear system-size dependence for lattice Hamiltonians but suffer from polynomial accuracy scaling, and post-Trotter methods, which achieve logarithmic accuracy dependence but lose the favorable locality structure. The physical intuition that one should “let Trotter handle the

majority of the simulation, with LCU completing the remainder” is sound and represents a natural interpolation between these regimes.

The treatment of the Trotter remainder $VK(x) = U(x)SK(x)$ and its Taylor series expansion is mathematically rigorous. The authors correctly exploit the order conditions established in prior work, particularly the vanishing of expansion terms $FK,1$ through FK,K , which follows from the well-known property that Trotter formulas match the exact evolution to order K in the time step. The observation that the leading-order terms $FK,K+1$ through $FK,2K+1$ are anti-Hermitian is crucial for the pairing construction, and the proof via the BCH formula in Equation 69 is correct. However, I have concerns about the treatment of lattice Hamiltonians in Section V. The authors define $H = A + B$ where A contains even-indexed terms and B contains odd-indexed terms, but the boundary conditions are not clearly specified. For periodic boundary conditions, the counting arguments in Proposition 7 need modification, and the “light cone” analysis in Figure 6 would need to wrap around. The authors should clarify whether they assume open boundary conditions throughout.

The nested commutator bounds in Proposition 12 appear correct, but the constants involved deserve scrutiny. The factor $= 2(4 + 5)$, where $$ is the Trotter stage number, grows as $2(4 \oplus 2 \oplus 5^{(k-1)} + 5)$ for $2k$ -th order formulas, which becomes substantial for higher orders. While the authors focus on first- and second-order cases in their numerics, the scaling of these prefactors matters for practical applicability. The treatment of the Heisenberg model in Equation 20 as a test case is appropriate, with the decomposition in Equation 21 ensuring $= 4$ for uniform sampling. The padding technique introduced in Equations 17-19 to ensure uniform 1-norms across elementary commutators is clever and preserves the locality structure, but the virtual qubit construction in Figure 6 needs more careful justification. Specifically, when the light cone reaches the boundary after s adjoint operations, adding virtual qubits with sI operations maintains the algebraic structure but may introduce subtle issues in the sampling probability normalization that are not fully addressed.

The random sampling implementation in Section II.D is physically sound, with the Hadamard test circuit in Figure 9 providing an unbiased estimator of $\text{Tr}(OVV)$. The overhead in the sample complexity (Proposition 1) is correctly identified as the key cost of the random sampling approach, and maintaining $$ as a constant is essential for efficiency. The authors’ claim that the algorithm achieves “2 orders of magnitude smaller” gate counts than fourth-order Trotter for generic Hamiltonians and “3 to 4 orders of magnitude higher accuracy” for lattice systems with the same gate costs requires verification against specific parameter regimes. The numerical results in Figure 8 support these claims, but the comparison methodology deserves scrutiny: the fourth-order Trotter bounds are analytical (from Ref. [12]) while the PTSC results appear to be computed from the LCU formula construction. This asymmetry could favor the Trotter-LCU approach. I recommend the authors provide a comparison using both analytical bounds and empirical circuit depths for all methods.

Questions for Authors: First, can you clarify the boundary conditions assumed for the lattice Hamiltonian analysis, and if periodic boundaries are intended, how do the light-cone arguments in Section V.B need to be modified? Second, for the Heisenberg model results in Figure 8(c), what is the specific procedure for computing the “fourth-order Trotter commutator bound” and can you confirm it uses the nested commutator form from Proposition M.1 of Ref. [20] rather than cruder estimates? Third, the virtual qubit padding construction appears to introduce extra terms in the LCU formula even though they contribute zero to the final result; does this affect the variance of the random sampling estimator, and if so, is this accounted for in your sample complexity analysis?

Verdict: 7/10. Recommendation: minor-revisions.

3.2.2 Voice 2 — Reviewer 2 (Algorithmic novelty)

The central algorithmic contribution of this work is the observation that LCU compensation of Trotter remainders can achieve better-than- $2K$ -th order scaling using only K -th order Trotter formulas. This is a genuine insight that has not appeared in the literature in precisely this form. However, the landscape of Hamiltonian simulation has evolved rapidly, and the novelty must be assessed against recent developments. The random sampling implementation of LCU, which the authors credit to Refs. [30, 33-35], has been substantially developed in works by Wan, Berta, and Campbell (Ref. [34], PRL 2022) and Faehrmann et al. (Ref. [35], Quantum 2022). The specific contribution here is the construction of LCU formulas that exploit Trotter order conditions and the pairing technique using Euler’s formula, neither of which appears in prior random-LCU work. This constitutes a novel algorithmic element.

The comparison to the most recent literature reveals both strengths and gaps. The work by Childs, Ostrander, and Su on randomized Trotter (Ref. [12], Quantum 2019) achieves improved practical performance

through random permutation of Trotter stages, but maintains the polynomial accuracy scaling. The authors correctly identify that their approach achieves time scaling of $1 + 1/(2K+1)$ compared to $1 + 1/K$ for standard K -th order Trotter, and logarithmic accuracy scaling for PTSC. However, I note that very recent work by Cho, Berry, and Hsieh (Ref. [47], PRA 2024, cited in the manuscript) considers “similar ideas to compensate the Trotter error using randomized unitary operators.” The authors acknowledge this in Section VI but do not provide a detailed comparison. This is a significant oversight given that Ref. [47] appeared on arXiv in the same timeframe as earlier versions of this manuscript. A more thorough comparison establishing the precise relationship and relative advantages would strengthen the novelty claim.

The algorithmic structure involves several components whose individual novelty varies. The Taylor series expansion of Trotter remainders is straightforward, and the observation that leading-order terms are anti-Hermitian follows from standard BCH analysis. The true novelty lies in the pairing construction (Equations 45 and 75) and the nested commutator compensation for lattice Hamiltonians. The pairing technique essentially converts additive error terms of the form $I + iV$ into rotation unitaries $\exp(iV)$, reducing the 1-norm from $1 + \epsilon$ to $(1 + \epsilon)^2$. This is elegant but also somewhat obvious once stated; the key question is whether it constitutes a “strict domination” improvement over prior work. Examining the complexity claims in Theorem 1 and Theorem 2, we have gate complexity $O((t)^{1+1/(2K+1)}(KL + \log(1)/\log\log(1)))$ for PTSC and $O(n^{1+2/(2K+1)} t^{1+1/(2K+1)} \epsilon^{-1/(2K+1)})$ for NCC on lattice Hamiltonians. These should be compared against the best known results: $O((t)^{1+1/K} \epsilon^{-1/K})$ for K -th order Trotter, $O(n \epsilon t \log(1))$ for post-Trotter methods on generic Hamiltonians, and $O(n^{1+1/K} t^{1+1/K} \epsilon^{-1/K})$ for K -th order Trotter on lattice Hamiltonians. The PTSC algorithm achieves a strict Pareto improvement in the (t, ϵ) parameter space for fixed system size, and NCC achieves a Pareto improvement in the (n, t, ϵ) space for lattice systems. These are legitimate improvements, though the practical margins depend strongly on prefactors.

The claims about orders of magnitude improvements (2 orders for generic Hamiltonians, 3-4 orders of accuracy for lattice systems) require careful verification. From Figure 8(a), at $t = 10\epsilon$ and $\epsilon = 10$, the fourth-order Trotter gate count is approximately 10^2 while fourth-order PTSC is approximately 10 , yielding a ratio of about 100ϵ , consistent with the “2 orders of magnitude” claim. From Figure 8(c), at the same gate count (approximately 10 Rz gates), fourth-order Trotter achieves 10ϵ while second-order NCC achieves 10 , a ratio of 10 in accuracy, consistent with the “3-4 orders of magnitude” claim. These ratios are indeed achievable but occur in specific parameter regimes; the claims should be qualified accordingly. The authors should also note that the comparison uses analytical bounds for Trotter, which are known to be loose by constant factors in practice.

Questions for Authors: First, can you provide a detailed comparison with Ref. [47] (Cho, Berry, Hsieh) establishing precisely which algorithmic elements are shared and which are novel to your work? Second, for the claimed “2 orders of magnitude” improvement, can you compute the gate counts using empirical Trotter error estimates (e.g., from numerical simulation) rather than analytical bounds to provide a more balanced comparison? Third, the transition from PTSC to NCC occurs when locality structure becomes important; can you characterize more precisely the crossover point in the (n, L, t, ϵ) parameter space where NCC becomes preferable to PTSC? Fourth, have you considered combining your approach with randomized Trotter (Ref. [12]) to potentially achieve further improvements?

Verdict: 6/10. Recommendation: major-revisions.

3.2.3 Voice 3 — Reviewer 3 (Empirical evidence)

The numerical evidence presented in this manuscript centers on gate count comparisons in Figure 8 and the resource estimation framework described in Section II.E and Appendix G. While the theoretical complexity results are rigorously derived, the empirical validation raises several methodological concerns that must be addressed before publication. The gate counting procedure is described as follows: circuits are compiled to CNOT gates, single-qubit Clifford gates, and non-Clifford $Rz()$ gates, with T-gate counts derived using the compilation formula from Ref. [54] with $cT \approx 66$ T gates per Rz gate at precision 10^{-2} . This is a reasonable procedure, but the manuscript does not provide confidence intervals or variance estimates for any of the reported gate counts. Given that the PTSC and NCC algorithms involve random sampling, the gate counts are random variables whose distributions should be characterized.

For the random sampling implementation, the sample complexity is $O(1/\epsilon)$ where ϵ is the LCU 1-norm and ϵ is the estimation precision (Proposition 1). The authors set $\epsilon = 2$ for the comparisons in Figure 8, which implies a 16ϵ overhead in sample number. However, the plots show single-shot gate counts rather than total resource requirements including the sampling overhead. This is explicitly acknowledged (“the random-sampling Trotter-

LCU algorithm will require an additional factor of 16 in the sample number”), but it would be more informative to present total gate counts including the sampling repetitions required to achieve a target observable estimation precision. Furthermore, the claim that PTSC achieves “2 orders of magnitude smaller” gate counts assumes that the fourth-order Trotter analytical bounds are tight. The authors use bounds from Ref. [12] for randomized Trotter, but it is well-known that analytical Trotter error bounds can overestimate the true error by orders of magnitude. A more rigorous comparison would include numerical estimates of the actual Trotter error for specific Hamiltonians.

The ablation study structure is incomplete. The manuscript presents results for 0th-order PTSC, 2nd-order PTSC, 4th-order PTSC, and 2nd-order NCC, but does not systematically vary other algorithmic choices. Key questions include: what is the effect of the truncation order sc on practical performance, how does the pairing technique compare to direct compensation without pairing, and what is the sensitivity to the padding procedure in the NCC algorithm? The authors mention in Section VI that “we can think about a hybrid implementation of different algorithms” but do not provide any numerical evidence for such hybrid approaches. A complete ablation would also include the effect of measurement and reset versus maintaining the ancilla qubit coherently (as discussed in Figure 10), which affects the circuit depth and connectivity requirements differently.

The absence of a systematic failure modes analysis is concerning. The manuscript does not discuss regimes where the Trotter-LCU approach may perform poorly. For instance, when the LCU 1-norm becomes large (which occurs when the leading-order Trotter error terms have substantial norm), the sample complexity grows as $\frac{1}{\epsilon}$. The authors state that $\frac{1}{\epsilon}$ can be kept constant by increasing the segment number s , but this trades off against circuit depth. There should be a discussion of cases where this tradeoff becomes unfavorable. Additionally, the numerical results focus on specific model Hamiltonians (the 2-local Hamiltonian in Equation G2 and the Heisenberg model in Equation 20), but no systematic study of Hamiltonian families is provided. The claim that the approach is “generic for L-sparse Hamiltonians” should be substantiated with results across a broader class of systems. Finally, there is no mention of an audit trail or reproducibility infrastructure. The manuscript does not reference any code repository, supplementary data files, or scripts that would allow independent verification of the numerical claims. For a paper making quantitative resource estimation claims, this is a significant deficiency.

Questions for Authors: First, can you provide variance estimates or confidence intervals for the gate counts shown in Figure 8, accounting for the randomness in the sampling procedure? Second, can you include numerical estimates of actual Trotter error (not just analytical bounds) for the Hamiltonians studied, to provide a more balanced comparison with your method? Third, what is the performance of your algorithms on quantum chemistry Hamiltonians, which you mention as a potential application but do not study numerically? Fourth, will you provide a code repository with scripts to reproduce all numerical results, including the gate count calculations and the LCU formula construction?

Verdict: 5/10. Recommendation: major-revisions.

3.2.4 Voice 4 — Devil’s Advocate

This manuscript, despite its mathematical sophistication, fails to deliver on its central promise of providing practical algorithms for near-term Hamiltonian simulation. Let me systematically dismantle the claims.

The foundational premise that these algorithms are “easy to implement” is misleading at best. The authors repeatedly contrast their approach with “post-Trotter” methods that require “multicontrolled Toffoli gates” and “many ancillary qubits.” But examining their own construction reveals substantial hidden complexity. The PTSC algorithm requires sampling from a distribution over Taylor expansion orders up to sc , followed by sampling Pauli strings based on multinomial distributions with $O(L)$ outcomes per trial. For each sampled configuration, one must compute the Pauli rotation angle $\theta = \tan^{-1}(\dots)$ where θ depends on the full expansion structure. The NCC algorithm is worse: it requires traversing “light cone” regions in the commutator structure, sampling adjoint operators with carefully computed probabilities, and handling boundary padding with virtual qubits. Algorithm 2 in Appendix D spans two pages of pseudocode. Calling this “easy to implement” when compared to, say, a straightforward fourth-order Trotter circuit that requires no classical preprocessing is disingenuous.

The claimed resource advantages evaporate under closer scrutiny. The authors proudly state “2 orders of magnitude smaller” gate counts, but this comparison is fundamentally unfair. They compare their algorithms against the analytical worst-case bound for fourth-order Trotter from Ref. [12], which is known to overestimate the actual error by substantial margins. In contrast, the Trotter-LCU gate counts appear to be computed from the explicit LCU formula construction, essentially a best-case estimate. A fair comparison would use empirical error estimates for both methods. Furthermore, the random sampling implementation incurs a $\frac{1}{\epsilon}$ overhead in

sample complexity. With $\epsilon = 2$, this is a factor of 16 in the number of circuit repetitions. The authors bury this fact, mentioning it only in passing, but it means that the “2 orders of magnitude” reduction in per-circuit gate count translates to at most a factor of $100/16 \approx 6.25$ reduction in total gate resources. This is a much more modest improvement that may not justify the algorithmic complexity.

The nested commutator compensation algorithm, presented as enabling near-optimal system-size scaling for lattice Hamiltonians, has severe limitations that are not adequately discussed. The performance guarantee in Theorem 2 applies only to one-dimensional lattice Hamiltonians of the form $H = \sum H_i$. For two-dimensional lattices, which are far more relevant for condensed matter physics and quantum chemistry, the “light cone” structure changes dramatically, and the $O(n)$ commutator bounds become $O(n^d)$ for d -dimensional systems. The authors never explicitly state this limitation. Additionally, the NCC algorithm only compensates the leading-order terms (order $K+1$ through $2K+1$), leaving a residual error of $O(x^{-(2K+2)})$. This means the accuracy scaling is $\sim (-1/(2K+1))$, which is polynomial rather than logarithmic. The PTSC algorithm achieves logarithmic accuracy scaling but loses the favorable system-size dependence. The authors present this as a choice between two advantageous regimes, but in reality, neither algorithm achieves both simultaneously, which is the actual holy grail of Hamiltonian simulation.

The numerical results are cherry-picked to favor the proposed methods. Figure 8(a) shows time dependence at fixed $n = 20$ and $t = 10$, a regime deliberately chosen to showcase the superior time scaling of PTSC. At shorter times ($t < 10$), the methods are comparable. Figure 8(c) uses $n = t$ for the lattice Hamiltonian, which obscures the separate contributions of system size and evolution time. Most damningly, the authors do not compare against quantum signal processing (QSP) in the main text results for lattice Hamiltonians, despite QSP being the state-of-the-art for generic Hamiltonians. The comparison with QSP in Appendix G shows that PTSC has higher T-gate counts than QSP (Figure 14), completely undermining the “easy to implement” narrative. On fault-tolerant quantum computers where T gates dominate the cost, QSP remains superior. The authors try to salvage this by noting PTSC has lower CNOT counts and qubit requirements, but these are secondary considerations for fault-tolerant architectures.

This paper represents solid theoretical work that has been oversold as a practical breakthrough. The algorithms are mathematically elegant but algorithmically complex, the resource advantages are overstated through unfair comparisons, and the fundamental limitations are underemphasized. I recommend rejection unless the authors substantially revise their claims and provide balanced comparisons using empirical error estimates for all methods.

Recommendation: reject.

3.2.5 Voice 5 — Editor-in-Chief synthesis

Having reviewed all four assessments, I find substantial merit in this manuscript alongside significant concerns that must be addressed. Let me reconcile the conflicting perspectives.

Reviewer 1 raises valid technical points about boundary conditions and the asymmetry in comparison methodologies but ultimately finds the physics correct and the core construction sound. Reviewer 2 acknowledges genuine algorithmic novelty in the pairing technique and nested commutator compensation but correctly identifies the missing comparison with concurrent work (Ref. [47]) as a significant gap. Reviewer 3’s concerns about empirical methodology are well-founded: the lack of confidence intervals, incomplete ablation studies, and absence of reproducibility infrastructure are problematic for a paper making quantitative resource claims. The Devil’s Advocate raises the most serious challenges: the “easy to implement” framing is oversold, the comparison methodology favors the proposed methods, and the practical advantage after accounting for sampling overhead is much smaller than advertised.

On the question of novelty, I side largely with Reviewer 2’s assessment. The pairing construction using Euler’s formula and the nested commutator compensation for lattice Hamiltonians represent genuine contributions that extend beyond straightforward combinations of known techniques. However, the relationship to Ref. [47] must be clarified explicitly. The authors acknowledge this work in Section VI but dismiss it with the phrase “similar ideas”; this is insufficient for a venue like PRX Quantum where establishing priority and precise technical distinctions is essential.

The Devil’s Advocate’s critique of the comparison methodology is largely valid but somewhat overstated. The authors do use analytical bounds for Trotter, which are conservative, but this is standard practice in the literature, and the same bounds are used in papers like Ref. [12] and Ref. [20]. A fairer criticism is that the sampling overhead should be more prominently displayed. The authors mention this in Section III.B, but

the gate count figures in Section II.E do not include this overhead, which could mislead readers about total resource requirements. That said, the Devil’s Advocate’s claim that the advantage reduces to “approximately $6\mathcal{E}$ ” incorrectly conflates gate counts with sample counts; the total gate count is still reduced even accounting for increased sampling, just not by two orders of magnitude.

For PRX Quantum, which requires both significant scientific advance and thorough technical presentation, this manuscript falls short in its current form but can be elevated with targeted revisions. The core ideas are sound and represent a meaningful contribution to the Hamiltonian simulation literature. The main deficiencies are in presentation and benchmarking methodology rather than in the underlying science.

Must-fix items before resubmission, ordered by severity:

1. Provide an explicit, detailed comparison with Ref. [47] (Cho, Berry, Hsieh), establishing precisely which algorithmic elements are shared, which are novel, and quantifying relative performance where applicable.
2. Revise the resource comparison methodology to include either (a) empirical Trotter error estimates for the test Hamiltonians to match the precision of the LCU formula construction, or (b) analytical bounds for both methods computed using the same framework.
3. Present total resource counts including sampling overhead prominently in the main text, either as separate figures or as multipliers applied to the single-shot counts.
4. Add a Failure Modes section discussing regimes where the approach performs poorly, including quantitative guidance on the $(n, L, t,)$ parameter crossovers.
5. Clarify boundary condition assumptions throughout the lattice Hamiltonian analysis and state explicitly that the NCC complexity bounds apply to one-dimensional systems.
6. Provide a reproducibility package (code repository or supplementary materials) containing the scripts and data files necessary to regenerate all numerical results.
7. Add variance or confidence interval estimates to the gate count results, or justify why the quantities reported are deterministic.

3.2.6 Vote table

Voice	Recommendation	Confidence 1-10
Reviewer 1	minor-revisions	7
Reviewer 2	major-revisions	6
Reviewer 3	major-revisions	7
Devil’s Advocate	reject	8
Editor-in-Chief	major-revisions	8

3.3 QN — Logical-fallacy report

Model: `claude-opus-4-5-20251101` · 11 quantum-CS-specific + standard. QN-specific stage.

Finding 1

- **Fallacy:** cherry-picked-baseline
- **Severity:** medium
- **Location:** Section II.E, Table I, and throughout the numerical comparisons
- **Evidence:** “For a fair comparison, we set the 1-norm of all LCU formulas \dagger to be constant. . . We compile their quantum circuits to CNOT gates, single-qubit Clifford gates, and single-qubit Z-axis rotation gates $Rz() = e^{iZ}$. Here, we mainly compare the number of $Rz()$ gates since they are the most resource-consuming part on a fault-tolerant quantum computer.”

- **Why it’s the fallacy:** The paper compares primarily against fourth-order Trotter with analytical bounds from Ref. [12] and [20], but does not compare against other recent hybrid or randomized Trotter methods that may offer competitive performance. The comparison with QSP in the appendix uses a specific qROM implementation from Ref. [52], but other block-encoding schemes or more recent QSP variants (e.g., quantum eigenvalue transformation with different ancilla strategies) are not considered. The baseline selection favors scenarios where the proposed method excels (high accuracy requirements, moderate system sizes).
 - **Suggested fix:** Include comparisons with additional state-of-the-art methods such as higher-order randomized Trotter methods from Ref. [47] (Cho et al., cited in the paper itself as concurrent work), and discuss why certain baselines were omitted. Acknowledge that different methods may be preferable in different parameter regimes.
-

Finding 2

- **Fallacy:** asymptotic-only-claim
 - **Severity:** medium
 - **Location:** Theorem 1 and Theorem 2 statements, Section IV.C and Section V.C
 - **Evidence:** “From Theorem 1 we can see that, by introducing the LCU compensation, the time scaling of the Kth-order Trotter-LCU algorithm improve the bare Kth-order Trotter time scaling from $1 + 1/K$ to $1 + 1/(2K + 1)$. Moreover, the accuracy scaling is exponentially improved.”
 - **Why it’s the fallacy:** The theoretical claims focus heavily on asymptotic scaling improvements (e.g., time scaling exponents, accuracy scaling), but the numerical demonstrations are limited to relatively small system sizes ($n = 12, 20, 50$) and modest simulation times. The asymptotic regime where the improved exponents dominate over constant factors is not clearly demonstrated. The paper acknowledges practical constants (e.g., the \dagger sampling overhead from Proposition 1) but does not systematically show when the crossover point occurs where the new method outperforms baselines.
 - **Suggested fix:** Add explicit crossover analysis showing at what system size, time, or accuracy threshold the asymptotic improvements begin to dominate. Provide a table or figure showing the regime of practical advantage versus the regime where constant factors dominate.
-

Finding 3

- **Fallacy:** conflated-regimes
 - **Severity:** medium
 - **Location:** Section II.E, Figure 8, and Section VI Conclusion
 - **Evidence:** “From Fig. 8(c), we can see that, while enjoying near-optimal system-size scaling similar to the fourth-order Trotter algorithm which is currently the best one for lattice Hamiltonians, the second-order NCC algorithm shows better accuracy dependence than fourth-order Trotter algorithm. Particularly, using the same gate number as the fourth-order Trotter, we are able to achieve a 3 to 4 orders of magnitude higher accuracy .”
 - **Why it’s the fallacy:** The numerical results shown in Figure 8(c) are for $n = 12$ and $n = 50$ qubit Heisenberg chains, which are relatively small systems. The extrapolation to larger, more practically relevant system sizes (hundreds or thousands of qubits for fault-tolerant quantum computing) is implicit but not demonstrated. The claim of “3 to 4 orders of magnitude higher accuracy” is based on these small-system numerics, and it’s unclear whether this advantage persists at larger scales where the nested commutator structure may become more complex.
 - **Suggested fix:** Explicitly state that the numerical demonstrations are limited to small-to-moderate system sizes and discuss how the advantage may scale. Consider adding extrapolation analysis or explicitly noting that the 3-4 orders of magnitude improvement is demonstrated only for the tested system sizes.
-

Finding 4

- **Fallacy:** hasty-generalization
 - **Severity:** medium
 - **Location:** Section VI Conclusion, and Section II.E
 - **Evidence:** “These algorithms provide an easy-to-implement approach to achieve a low-cost and high-precision Hamiltonian simulation.”
 - **Why it’s the fallacy:** The claim of “easy-to-implement” is supported primarily by the fact that only one ancilla qubit is needed and Pauli rotation gates are used. However, the classical sampling procedure (Algorithm 1, Algorithm 2, Fig. 5, Fig. 7) involves non-trivial multinomial sampling, light-cone tracking, and padding procedures. The paper demonstrates the algorithm primarily on idealized lattice Hamiltonians (Heisenberg model) and 2-local generic Hamiltonians. Generalization to more complex Hamiltonians (e.g., quantum chemistry with many-body terms, non-local interactions) is discussed briefly in Appendix E but without numerical validation.
 - **Suggested fix:** Temper the “easy-to-implement” claim by explicitly discussing the classical preprocessing complexity, or provide numerical demonstrations for more complex Hamiltonian structures beyond the idealized test cases.
-

Finding 5

- **Fallacy:** pareto-cherry-picked-axes
 - **Severity:** medium
 - **Location:** Table I, Figures 8, 13-17, and Section II.E
 - **Evidence:** “In Table I, we compare the implementation complexity and the gate complexity in a single round of experiment of the 0th-order PTSC, Kth-order PTSC, and Kth-order NCC algorithms to previous Hamiltonian simulation algorithms.”
 - **Why it’s the fallacy:** The comparison metrics emphasize gate complexity (specifically Rz and CNOT counts) and accuracy scaling, but downplay or omit other important dimensions: (1) the t sampling overhead is mentioned but not factored into the total cost comparisons in Table I; (2) classical preprocessing time for sampling is not compared; (3) qubit count comparison is shown only in Fig. 15 for QSP but not systematically tabulated; (4) the “implementation hardness” column in Table I uses qualitative labels (“Easy” vs “Hard”) without precise definitions. The axes chosen for comparison favor the proposed methods.
 - **Suggested fix:** Include the t sampling overhead explicitly in the complexity comparisons when relevant, or present a more comprehensive resource table that includes total sample complexity, classical computation overhead, and ancilla requirements on equal footing.
-

3.4 QN — Stage-6 CQE narrative

Model: claude-opus-4-5-20251101 · 6-dim Collaboration Quality Evaluation. QN-specific.

Process Summary: QuantumNovelty Run Evaluation

3.4.1 Composite Verdict

The run achieved a **composite score of 23 out of 100**, calculated via geometric mean across six dimensions. Per standard interpretation scales, this places the collaboration firmly in the “**Critical Deficiencies**” tier (scores 1-30), indicating that the run failed to produce the fundamental artifacts required for a credible scientific workflow. A geometric mean this low signals that no single dimension performed well enough to compensate for the others—the run exhibits systemic gaps rather than isolated weaknesses.

To be direct: a score of 23 represents a workflow that generated minimal verifiable output. The geometric mean is unforgiving by design; it penalizes runs where any dimension collapses. Here, nearly every dimension collapsed.

3.4.2 Strongest Dimension: Communication (Score: 40)

The highest-scoring dimension was **Communication** at 40 points, though this warrants careful interpretation. Both probes—“logical fallacies absent” and “reviewer panel verdict”—scored 40, but the evidence reveals these scores reflect *absence of evaluation* rather than *positive verification*. The “logical_fallacies skill not run” and “no review_panel.md found” entries indicate these checks were never executed.

A score of 40 achieved through non-execution is not a strength—it’s an artifact of the scoring system assigning default values when probes cannot find contrary evidence. The run produced so little written output that there was nothing to evaluate for fallacies or review. This “highest” score therefore tells us something damning: the workflow never progressed far enough to generate prose that could be assessed for communication quality.

3.4.3 Weakest Dimension: Novelty Rigour (Score: 8)

The weakest dimension was **Novelty Rigour** at 8 points, representing a near-total failure of the discovery validation pipeline. Two probes contributed to this collapse:

1. “**augmented baseline catalog present**” scored 10/100, with evidence showing “baseline_catalog has 0 rows.” This means the run operated without any reference corpus of known results. Without a baseline catalog, there is no principled way to distinguish rediscovery from genuine novelty. The system was flying blind.
2. “**strict-domination comparator run**” scored 5/100, with “novelty_verdict.json not found.” The comparator—the mechanism that determines whether candidate solutions strictly dominate known baselines—never executed or never produced output.

The stage responsible for this failure is clearly the **initialization and catalog-building phase**. A baseline catalog with zero rows indicates the workflow either skipped the literature/prior-art ingestion step entirely or encountered a silent failure during catalog population. Everything downstream—novelty comparison, Pareto classification, claim generation—becomes meaningless without this foundation.

A Novelty Rigour score of 8 means any “discoveries” from this run carry zero epistemic weight. They cannot be distinguished from well-known results.

3.4.4 Three Highest-Leverage Improvements

1. Populate the Baseline Catalog Before Any Optimization Runs The single highest-leverage fix is ensuring the baseline catalog contains substantive prior-art data before the main workflow begins. The “augmented baseline catalog present” probe’s evidence (“baseline_catalog has 0 rows”) indicates this step was skipped or failed silently. Implementation should include: - A blocking check that halts the workflow if baseline_catalog row count is below a configurable threshold - Automated ingestion from at least two sources (literature database + prior run archives) - Explicit logging confirming catalog population counts

Without this, the entire novelty-detection apparatus is inert.

2. Emit Core Verification Artifacts as Mandatory Checkpoints Multiple probes failed because expected files simply don’t exist: `novelty_verdict.json`, `audit_claims.py`, `paper.tex`, `wilson_annotations.md`, `ablation_results.json`, `ratio_recompute.md`, `review_panel.md`. The workflow should treat these as **mandatory checkpoints** rather than optional outputs. Specifically: - Make `novelty_verdict.json` emission a hard requirement before any claims can be registered - Gate the “results finalization” stage on existence of `audit_claims.py` - Require Wilson confidence interval computation for any quantitative claim

The current run demonstrates what happens when artifact generation is advisory: nothing gets generated.

3. Enforce Cross-Vendor Execution for Falsifiability The “cross-LLM with multiple vendors” probe scored 40 with evidence “vendors used: []”—an empty list. For any LLM-in-the-loop scientific workflow, single-vendor execution cannot establish that results stem from the method rather than model-specific artifacts. The next run should: - Require configuration of at least two LLM vendors before workflow start - Run identical prompts through each vendor and log response divergence - Surface any vendor-specific anomalies in a dedicated comparison artifact

This addresses Falsifiability while also strengthening Methodological Rigour through implicit ablation.

3.4.5 Conclusion

This run produced a scaffold with no content. The score of 23 accurately reflects a workflow that initialized but never executed its core verification stages. The path forward requires treating artifact generation as blocking rather than optional, starting with the baseline catalog that anchors all novelty claims.

4 ARS academic-paper-reviewer — full per-agent prose

Every subsection below is ARS output (`_run/ars/`). Each agent is one independent LLM call; the editorial synthesizer (Phase 2) sees the 5 reviewer cards in its context, but the 5 reviewers are blind to each other.

4.1 ARS — Phase 0: Field analyst

Model: `claude-haiku-4-5-20251001` · Identifies the paper’s field, dynamically configures 5 reviewer personas.

Field Analysis Report

4.1.1 Paper Basic Information

- **Title:** Simple and high-precision Hamiltonian simulation by compensating Trotter error with linear combination of unitary operations
- **Abstract length:** ~200 words
- **Full text length:** ~12,000 words (partial text provided, appears to be a substantial theoretical physics/quantum computing paper)
- **Number of references:** ~40+ (based on citations visible in text)

4.1.2 Field Analysis

Dimension	Analysis Result
Primary Discipline	Quantum Computing / Quantum Information Science
Secondary Disciplines	Mathematical Physics, Computational Complexity Theory, Numerical Analysis
Research Paradigm	Theoretical/Conceptual Analysis with Algorithmic Development
Methodology Type	Mathematical Proof + Algorithm Design + Complexity Analysis
Target Journal Tier	Q1 — The paper targets PRX Quantum, a premier journal in quantum information; the work presents novel algorithmic contributions with rigorous proofs, builds on foundational work (Trotter, LCU methods), and demonstrates significant practical improvements (2-4 orders of magnitude in accuracy)
Paper Maturity	Pre-submission — Structure is complete with theorems, propositions, detailed proofs in appendices, numerical comparisons, and polished notation; minor refinements may be needed

4.1.3 Recommended Target Journals (Top 3)

1. **PRX Quantum** — Premier venue for quantum algorithms and quantum simulation; this paper’s combination of practical algorithm design with rigorous theoretical guarantees fits perfectly with the journal’s mission to publish “high-quality, high-impact research”
2. **Quantum** — Open-access high-impact journal for quantum science; alternative if PRX Quantum declines; strong fit for algorithmic contributions
3. **npj Quantum Information** — Nature portfolio journal covering quantum computing advances; the practical gate-count improvements demonstrated would appeal to this readership

4.1.4 Reviewer Configuration Cards

Reviewer Configuration Card #1 **Role:** Editor-in-Chief (Handling Editor) **Identity Description:** Senior Editor at *PRX Quantum*, former faculty at Caltech with expertise in quantum algorithms and fault-tolerant quantum computing. Has previously handled landmark papers on quantum signal processing and post-Trotter simulation methods. Serves on the steering committee for QIP (Quantum Information Processing) conference.

Review Focus: 1. Does this work represent a genuine advance over existing Trotter and post-Trotter methods, or is it an incremental combination of known techniques? 2. Is the claimed 2-4 orders of magnitude improvement in gate counts substantiated by fair comparisons with appropriate baselines? 3. Will this paper influence how the quantum computing community implements Hamiltonian simulation in practice?

Will particularly care about: Whether the paper clearly articulates why combining Trotter with LCU is non-trivial and what new insights enable this synthesis. Also concerned with whether the paper is accessible to the broad PRX Quantum readership beyond specialists in product formulas.

Possible blind spots: May underweight implementation challenges on actual quantum hardware; may not deeply scrutinize the numerical benchmarks against alternative methods like qubitization or quantum signal processing.

Reviewer Configuration Card #2 Role: Peer Reviewer 1 — Methodology (Mathematical Rigor & Proof Verification) **Identity Description:** Professor of Mathematics specializing in operator theory and matrix analysis, with extensive work on product formula error bounds (Magnus expansion, BCH formula). Has published in *Communications in Mathematical Physics* and *Journal of Mathematical Physics*. Known for meticulous proof verification and identifying subtle errors in asymptotic bounds.

Review Focus: 1. Verify the correctness of the key propositions (Propositions 3, 4, 6) and the derivation of the 1-norm bounds for the LCU formulas 2. Check whether the order-pairing technique using Euler’s formula (Equation 7) is applied correctly and whether the claimed 1-norm suppression (from $O(x)$ to $O(x)\delta$ to $O(x)$) is mathematically rigorous 3. Examine the Taylor series truncation error analysis and whether the bounds in Equations (50), (79), and (113) are tight or overly conservative

Will particularly care about: The transition from Equation (72) to (74) where anti-Hermitian structure is exploited—whether all Hermitian terms genuinely cancel. Also the validity of applying Proposition 2 (product of LCU formulas) in the random sampling context.

Possible blind spots: May not assess practical relevance or implementation feasibility; may not compare with alternative algorithmic approaches outside the Trotter/LCU paradigm.

Reviewer Configuration Card #3 Role: Peer Reviewer 2 — Domain Expert (Quantum Simulation Algorithms) **Identity Description:** Research scientist at a major quantum computing company (Google AI Quantum or IBM Quantum), specializing in Hamiltonian simulation algorithms for near-term and fault-tolerant quantum computers. Has published extensively on Trotter error bounds with commutator scaling, including work building on Childs et al. and Campbell’s tightened bounds. Regularly benchmarks algorithms for chemistry and condensed matter applications.

Review Focus: 1. Assess whether the comparison with fourth-order Trotter (Figures 8a-c) uses state-of-the-art bounds (particularly the commutator bounds from Ref. [20]) or outdated loose bounds 2. Evaluate whether the gate counting methodology is fair—particularly the claim that Rz gates are the dominant cost in fault-tolerant implementation and whether T-gate counts would tell a different story 3. Examine whether the nested-commutator compensation (NCC) algorithm’s system-size scaling $O(n^{\{1+2/(2K+1)\}})$ is genuinely near-optimal compared to known lower bounds

Will particularly care about: Whether the overhead in sample complexity (Proposition 1) undermines the claimed advantages when total resource cost (gates & samples) is considered. Also whether the algorithm can be extended to non-Pauli Hamiltonians or requires the specific Pauli decomposition structure.

Possible blind spots: May be biased toward algorithms already implemented at their company; may not appreciate novel theoretical techniques if they don’t immediately translate to hardware advantages.

Reviewer Configuration Card #4 Role: Peer Reviewer 3 — Cross-disciplinary/Practical (Quantum Error Correction & Fault-Tolerant Implementation) **Identity Description:** Associate Professor working on fault-tolerant quantum computing architectures, specializing in the interface between abstract quantum algorithms and physical implementation. Has published on magic state distillation costs, logical qubit overhead, and

resource estimation for quantum chemistry applications. Affiliated with a national lab (e.g., Sandia or ORNL) conducting end-to-end resource estimates.

Review Focus: 1. Evaluate whether the “easy implementation” claim (Table I) holds under realistic fault-tolerant constraints—does the random sampling introduce additional challenges for syndrome extraction or error propagation? 2. Assess the ancilla qubit requirements: the paper claims only 1 ancillary qubit is needed, but does this account for measurement-based uncomputation and potential magic state factories? 3. Examine whether the mid-circuit measurement and reset scheme (Figure 2c) is compatible with current fault-tolerant architectures (surface codes, color codes) without excessive overhead

Will particularly care about: Whether the controlled-Pauli rotation gates (Figure 11b) can be efficiently synthesized into Clifford+T, and whether the $= \tan^{-1}(y)$ angles appearing throughout require expensive rotation synthesis. Also concerned about the classical sampling overhead—can it be performed efficiently during quantum circuit execution?

Possible blind spots: May not deeply engage with the mathematical elegance of the construction; may apply overly strict near-term feasibility criteria to what is presented as a theoretical contribution.

4.1.5 Review Strategy Recommendations

- **Critical tension to monitor:** Reviewer 2 (domain expert) may view this as an incremental improvement over existing post-Trotter methods, while Reviewer 3 (implementation expert) may be skeptical of practical claims. The EIC should weigh whether the theoretical novelty of the “order-pairing” technique compensates for any practical limitations.
- **Complementarity across reviewers:** Reviewer 1 (mathematics) will verify internal consistency of proofs; Reviewer 2 (algorithms) will contextualize within the competitive landscape; Reviewer 3 (implementation) will stress-test practical claims. Together they cover theory, algorithmic positioning, and realizability.
- **Special attention needed:** The paper’s core contribution—that combining Trotter with targeted LCU error compensation achieves better scaling than either alone—requires all reviewers to verify that the comparison baselines are fair. Figures 8a-c are likely to be scrutinized heavily.
- **Potential weakness to probe:** The sampling overhead (hidden in Proposition 1) could be a dealbreaker if not properly addressed. Reviewers should request clarification on total resource cost, not just gate counts.
- **Language/presentation note:** The paper is well-structured but dense; Reviewer 3 in particular may request a clearer “practical implications” section summarizing when practitioners should prefer this method over alternatives.

4.2 ARS — Phase 1a: EIC

Model: claude-haiku-4-5-20251001 · Editor-in-Chief journal fit, originality, overall quality.

4.2.1 EIC Review Report

Reviewer Identity Editor-in-Chief of PRX Quantum, evaluating this manuscript on Hamiltonian simulation algorithms that combine Trotter formulas with linear combination of unitary (LCU) methods. PRX Quantum publishes high-impact research across quantum information science, quantum computing, and quantum technologies, with particular interest in algorithmic advances that bridge theoretical innovation with practical implementation.

Overall Recommendation **Minor Revision**

Confidence Score 4 (High confidence - Quantum algorithms and Hamiltonian simulation are core topics for PRX Quantum)

Summary Assessment This paper proposes composite Hamiltonian simulation algorithms that combine Trotter formulas with LCU compensation to achieve improved time and accuracy scaling. The core idea—using Trotter circuits for the bulk of the simulation while employing LCU to compensate the leading-order Trotter error—is elegant and addresses a genuine need in the field. The authors demonstrate that by adding simple Pauli-rotation gates after each Trotter segment, one can achieve time scaling of $t^{\wedge}(1+1/(2K+1))$ compared to $t^{\wedge}(1+1/K)$ for bare K th-order Trotter, with exponentially improved accuracy dependence (logarithmic in $1/$ for PTSC algorithms).

The manuscript presents two algorithms: Paired Taylor-Series Compensation (PTSC) for generic Hamiltonians and Nested-Commutator Compensation (NCC) for lattice Hamiltonians with locality structure. The numerical comparisons suggest 2-4 orders of magnitude improvement over fourth-order Trotter in gate counts for practical scenarios. The work sits at a valuable intersection of theoretical rigor and practical applicability—requiring only one ancillary qubit and avoiding the multi-controlled gates that make post-Trotter methods challenging to implement near-term.

The contribution is timely given the push toward early fault-tolerant quantum computing. However, the presentation could be streamlined, and certain claims about practical advantages require stronger empirical grounding beyond analytical bound comparisons.

Strengths (3-5 items)

1. **Elegant conceptual contribution:** The insight that Trotter errors can be efficiently compensated using LCU formulas—leveraging the order condition to eliminate leading terms and Euler’s formula to pair remaining anti-Hermitian terms—represents a genuinely novel synthesis of two established approaches. The “best of both worlds” framing (commutator scaling from Trotter + high accuracy from LCU) is well-motivated and clearly demonstrated.
2. **Practical implementation focus:** The algorithms require only one ancillary qubit and simple Pauli-rotation gates, making them significantly more implementable than coherent LCU or QSP methods that need extensive ancilla systems and multi-controlled Toffoli gates. The random-sampling implementation (Sec. II.D, Figs. 5 and 7) provides explicit, efficient classical sampling procedures.
3. **Strong analytical framework with dual algorithms:** The paper provides rigorous performance guarantees (Theorems 1 and 2) while offering two complementary algorithms—PTSC for generic Hamiltonians achieving $\log(1/)$ accuracy scaling, and NCC for lattice Hamiltonians achieving nearly optimal system-size scaling $O(n^{\wedge}(1+2/(2K+1)))$. This breadth addresses different use cases effectively.
4. **Compelling numerical evidence:** The gate-count comparisons in Figure 8 demonstrate substantial practical advantages—2 orders of magnitude improvement over fourth-order Trotter for generic Hamiltonians, and 3-4 orders of magnitude higher achievable accuracy for lattice models at fixed gate budget. The comparisons against state-of-the-art analytical bounds add credibility.
5. **Clear pedagogical presentation of key ideas:** Figures 3 and 4 effectively illustrate how the “pairing” idea progressively suppresses the 1-norm, making the core technical insight accessible even to readers less familiar with LCU methods.

Weaknesses (3-5 items)

1. **Analytical bounds comparison may overstate practical advantage:** The numerical comparisons (Fig. 8) pit the authors’ algorithms against analytical Trotter bounds from Refs. [12, 20], which are known to be loose. A fairer comparison would include empirical Trotter error measured through numerical simulation, or at minimum acknowledge that analytical bounds are pessimistic. The claimed “2-4 orders of magnitude” improvement may be smaller in practice. *Improvement: Include numerical simulations of actual Trotter errors for the benchmark Hamiltonians, or clearly caveat that comparisons are bound-to-bound rather than performance-to-performance.*
2. **Limited discussion of sampling overhead:** While the paper acknowledges the $t^{\wedge}4$ sampling overhead (Proposition 1), it somewhat downplays this cost by “setting t to be constant.” For PTSC, this means $t^{\wedge}4 = 16$ (with $t = 2$), but the impact on total runtime relative to coherent Trotter implementation deserves

more explicit discussion. Additionally, the variance of the random-sampling estimator and its impact on sample complexity for different observables is not thoroughly analyzed. *Improvement: Add a dedicated subsection quantifying when the random-sampling approach outperforms coherent Trotter, including explicit sample-complexity breakeven analysis.*

3. **NCC gate-count analysis incomplete for higher orders:** The paper primarily presents second-order NCC results (Fig. 8c), noting that “precise higher-order NCC gate count analysis” is left for future study. This is a notable gap given that the theoretical scaling improvements are more dramatic at higher K . *Improvement: At minimum, provide rough estimates or bounds for fourth-order NCC, or justify why second-order is the primary practical regime.*
4. **Presentation length and density:** At 23 pages (main text) plus extensive appendices, the manuscript is long and technically dense. Some material, particularly the detailed derivations in Sec. IV and V, could be condensed by moving intermediate steps to appendices. The paper would benefit from a more streamlined main text emphasizing key ideas and results. *Improvement: Target a main text closer to 15 pages, moving detailed proofs to supplementary material while maintaining rigor.*

Detailed Comments

Journal Fit The paper fits PRX Quantum’s scope extremely well. Hamiltonian simulation is a core topic in quantum algorithms research, and the work addresses both theoretical foundations and practical implementation considerations. The balance between analytical rigor and near-term applicability aligns with the journal’s mission to publish impactful quantum information science research. The readership—quantum computing theorists, quantum algorithm developers, and those working toward early fault-tolerant implementations—will find this directly relevant.

Originality The conceptual contribution is original: while Trotter and LCU methods are both well-established, their systematic combination to compensate Trotter error using LCU formulas with constant 1-norm is novel. The “pairing” technique using Euler’s formula to double the suppression order of the 1-norm remainder (Eq. 7 and Fig. 3c) is a clever technical contribution. The nested-commutator expansion maintaining locality structure (Sec. V) extends this to lattice systems in a non-trivial way. The work builds substantively on prior literature (Refs. [8, 13, 20, 25, 33-35]) while contributing genuinely new methodology.

Significance If the claimed performance holds in practice, this work could influence how near-term and early fault-tolerant Hamiltonian simulation is performed. The algorithms offer an attractive middle ground: easier to implement than full post-Trotter methods, yet achieving better scaling than bare Trotter. For quantum chemistry applications (where high precision is needed) and lattice model simulations (where system size scales are relevant), the improvements could be practically meaningful. The impact is more incremental than transformative—refinement rather than paradigm shift—but valuable for the field’s practical progress.

Structural Coherence The paper is well-structured overall: Section II provides an accessible summary before technical details, and Theorems 1-2 clearly state the main results. However, the flow from PTSC (Sec. IV) to NCC (Sec. V) could be smoother—the connection between these algorithms and when to prefer one over the other is not immediately clear. The duplicate “Department of Computer Science” in the author affiliations (affiliation 4) should be corrected.

Title & Abstract The title is descriptive and accurate. The abstract effectively conveys the key contributions: both algorithms, their advantages (improved time/accuracy scaling, easy implementation), and quantitative improvement claims. The phrase “high-precision Hamiltonian simulation” appropriately signals the paper’s focus.

Conclusion Section VI provides adequate conclusions and future directions. The acknowledgment that empirical performance verification remains future work (implicit in the analytical focus) could be made more explicit.

Questions for Authors

1. How do the PTSC and NCC algorithms perform when Trotter error is measured empirically (through numerical simulation) rather than compared against analytical bounds? Can you provide evidence that the improvement persists in realistic comparisons?
2. For random-sampling implementations, what is the crossover point in circuit depth or observable accuracy where the $t^4 = 16$ sampling overhead becomes disadvantageous compared to simply running more coherent Trotter steps?
3. The paper focuses on expectation value estimation. How do the algorithms perform for other common tasks, such as state preparation fidelity or sampling from the output distribution?
4. For NCC applied to non-homogeneous lattice Hamiltonians, the “padding” technique (Sec. II.D, Eq. 18) introduces additional terms. How much does this inflate the effective 1-norm compared to homogeneous cases?

Minor Issues

- Affiliation 4 lists “Department of Computer Science” twice—please correct.
- Equation (21) appears to have a coefficient of 4 that seems inconsistent with standard Heisenberg Hamiltonian conventions—please clarify or correct.
- Figure 8 legend could be clearer about which curves use commutator bounds versus non-commutator bounds.
- Some notation inconsistencies: ϵ is used both for LCU accuracy and simulation accuracy; clarify usage or use distinct symbols.
- The sentence “This allowss us to achieve optimal gate complexity” (p. 16) has a typo (“allowss”).

Recommendation to Peer Reviewers For the Methodology Reviewer: Please scrutinize the 1-norm bounds in Propositions 3-6 and the gate-complexity derivations in Theorem 1-2. Are the asymptotic scalings tight, or could they be improved? Verify that the sampling procedures (Figs. 5, 7, Algorithm 1) are computationally efficient as claimed.

For the Domain Reviewer: Assess whether the numerical comparisons in Fig. 8 represent fair benchmarking against state-of-the-art Trotter implementations. Evaluate whether practitioners would find the implementation requirements (one ancilla, Pauli rotations, classical sampling) genuinely easier than alternatives.

For the Devil’s Advocate Reviewer: Stress-test the practical significance claims. The t^4 sampling overhead and the comparison against loose analytical bounds are potential vulnerabilities. Push on whether the “2-4 orders of magnitude” improvements survive under more realistic assumptions.

4.3 ARS — Phase 1b: Methodology reviewer

Model: c1aude-opus-4-5-20251101 · Peer Reviewer 1 research design, statistical validity, reproducibility.

4.3.1 Contract Paraphrase

From the perspective of methodology rigor, this contract asks me to evaluate a quantum computing paper proposing hybrid Trotter-LCU algorithms for Hamiltonian simulation. The paper must demonstrate correct methods and sound reasoning. I must assess whether the mathematical derivations are valid, the algorithm constructions are sound, and the claimed complexity improvements are properly justified.

Novelty (NOVELTY): I will evaluate whether the proposed combination of Trotter formulas with LCU compensation represents a methodologically sound new approach. From a methodology perspective, novelty means the technical construction must be non-trivial and the claimed improvements must derive logically from the algorithmic innovations rather than from artifacts of analysis choices.

Correctness (CORRECT): This is central to my review. I must verify that theorems, propositions, and lemmas are correctly stated and that the proofs (or proof sketches in the main text with full proofs in appendices) are mathematically valid. The LCU formula constructions, error bounds, gate complexity analyses, and commutator scaling arguments must all be technically sound.

Clarity (CLARITY): From a methodology standpoint, clarity means the algorithmic constructions and performance analyses must be presented with sufficient precision that an expert could verify the claims. The paper should clearly specify assumptions, define notation consistently, and present results in a way that enables reproducibility of the theoretical analysis.

Significance (SIGNIF): I will assess whether the claimed improvements (2 orders of magnitude gate reduction, 3-4 orders of magnitude accuracy improvement) are methodologically substantiated. The comparison methodology with prior work must be fair, and the performance claims must follow rigorously from the analysis rather than from optimistic parameter choices.

4.3.2 Scoring Plan

D1: NOVELTY what_to_look_for: - Whether the combination of Trotter formulas with LCU error compensation is a genuine algorithmic innovation vs. a straightforward concatenation of known techniques - Whether the “order-pairing” technique using Euler’s formula (Eq. 7) to suppress 1-norm represents a novel insight - Whether the nested-commutator compensation (NCC) approach offers methodological novelty in exploiting Hamiltonian structure - Prior work acknowledgment: do the authors accurately characterize the novelty relative to existing Trotter methods [8-22] and LCU methods [24-29]?

what_triggers_block: - The core technique is a trivial application of known methods with no novel combination or insight - The paper misrepresents prior work to inflate novelty claims - The “improvements” come from using different assumptions or benchmarks rather than genuine algorithmic innovation

what_triggers_warn: - Novelty is incremental: the core ideas exist in prior work but are combined in a new way without deep new insight - The paper’s novelty claims are overstated relative to what the technical content actually delivers - Some key ingredients (e.g., order-pairing) are not novel but borrowed from other contexts without attribution

D2: CORRECT what_to_look_for: - Validity of Theorem 1 (PTSC gate complexity) and Theorem 2 (NCC gate complexity): do the stated bounds follow from the derivations? - Correctness of Propositions 3-7: are the LCU formula constructions, 1-norm bounds, and error bounds mathematically sound? - Validity of the recurrence relation in Eq. 97 and the derivation of expansion terms - Whether the order condition arguments (Lemma 1, Lemma 2, Proposition 5) are correctly applied - Whether the complexity comparisons in Table I are fair (same assumptions across methods) - Whether the gate count estimates in Figures 8(a,b,c) follow from the analytical bounds or involve unjustified approximations

what_triggers_block: - A central theorem or proposition contains a mathematical error that invalidates the main claims - The gate complexity bounds are derived under hidden assumptions that favor the proposed method - The comparison with fourth-order Trotter uses different error metrics or assumes different circuit models, making comparisons invalid - Proof sketches omit critical steps that cannot be filled in by a competent reader

what_triggers_warn: - Minor errors in non-central results that don’t affect main conclusions - Some bounds are stated without proof and deferred to appendices, but the main text doesn’t provide enough intuition - Numerical estimates in Section II.E use approximations whose validity isn’t fully justified - Edge cases or parameter regimes are not fully analyzed

D3: CLARITY what_to_look_for: - Whether the algorithm descriptions (Figures 5, 7, Algorithm 1) are precise enough to implement - Whether notation is consistent throughout (e.g., definitions of $V_K(x)$, $F_{K,s}(x)$, η_s) - Whether the relationship between main text claims and appendix proofs is clearly signposted - Whether the random-sampling implementation details in Section IV.C are sufficiently specified - Whether the hierarchical sampling procedure is explained with enough detail for reproducibility

what_triggers_block: - Critical algorithm steps are ambiguous or underspecified - Notation inconsistencies make it impossible to follow the derivations - Main theorems cite appendix results that are not clearly identified

what_triggers_warn: - The paper is dense and would benefit from more intuitive explanations - Some implementation details are relegated to appendices without sufficient main-text summary - Figures require additional explanation to be self-contained

D4: SIGNIF what_to_look_for: - Whether the “2 orders of magnitude” gate reduction claim (abstract, Section II.E) is justified by the analysis - Whether the “3 to 4 orders of magnitude” accuracy improvement (abstract, Figure 8c) is a fair comparison - Whether the complexity improvements translate to practical advantages for realistic problem sizes - Whether the implementation requirements (single ancilla qubit, simple gates) represent genuine practical advantages - Whether the random-sampling overhead (μ^4 factor in Proposition 1) undermines the claimed advantages

what_triggers_block: - The claimed improvements are based on comparing best-case for the new method vs. worst-case for prior work - The practical parameter regime where improvements hold is not explicitly characterized - The comparison ignores significant factors (e.g., random sampling overhead)

what_triggers_warn: - The improvements are real but context-dependent; the paper doesn’t adequately characterize the parameter regime - The significance claims in the abstract are stronger than what the detailed analysis supports - The comparison methodology has some limitations that are acknowledged but could be clearer

[CONTRACT-ACKNOWLEDGED]

4.4 ARS — Phase 1c: Domain reviewer

Model: claude-haiku-4-5-20251001 · Peer Reviewer 2 literature coverage, theoretical framework, contribution.

4.4.1 Domain Review Report (Peer Reviewer 2)

Reviewer Identity Senior researcher in quantum computing algorithms with expertise in Hamiltonian simulation, quantum error correction, and fault-tolerant quantum computing. Extensive publication record in product formula methods and linear combination of unitaries techniques.

Overall Recommendation Minor Revision

Confidence Score 4

Summary Assessment This paper presents a novel hybrid approach combining Trotter formulas with linear combination of unitaries (LCU) methods for Hamiltonian simulation. The core contribution—using LCU to compensate Trotter error rather than performing simulation entirely with either method—represents a genuine and meaningful advance in the field. The theoretical framework is sound, building appropriately on established foundations from Suzuki, Childs, and others. The paper successfully bridges two major algorithmic paradigms that have traditionally been treated as competing approaches.

The literature coverage is adequate but not comprehensive, particularly regarding recent developments in randomized product formulas and quantum signal processing alternatives. The theoretical analysis is rigorous, with proper use of BCH expansion, order conditions, and commutator scaling arguments. The claimed improvements in gate complexity—2 orders of magnitude for generic Hamiltonians and 3-4 orders of magnitude in accuracy for lattice systems—are supported by the analytical bounds presented, though the numerical comparisons rely on analytical bounds rather than empirical benchmarks.

The paper makes appropriate use of the nested commutator framework from Childs et al. (2021) and extends it meaningfully. However, some theoretical nuances regarding the interplay between random sampling overhead and deterministic implementations deserve clearer treatment.

Strengths

1. **Novel Algorithmic Paradigm:** The fundamental insight that Trotter and LCU methods can be composed rather than competing is conceptually valuable. The “compensation” perspective—using LCU only for the residual error after Trotter—is elegant and leads to practical improvements.
2. **Rigorous Mathematical Framework:** The paper provides careful proofs for the 1-norm bounds (Propositions 3, 4, 6), order conditions (Lemma 1, Proposition 5), and gate complexity (Theorems 1, 2). The use of Euler’s formula for order-pairing (Eq. 7, 45) to reduce the 1-norm is technically sophisticated.
3. **Commutator Scaling Preservation:** The NCC algorithm (Sec. V) successfully maintains the commutator scaling advantages of Trotter methods while achieving improved accuracy dependence. This addresses a key limitation of standard post-Trotter methods for lattice Hamiltonians.
4. **Practical Implementation Path:** The random-sampling implementation (Figs. 5, 7) with single-ancilla circuits (Fig. 1b, 2b,c) offers a realistic near-term implementation pathway. The hierarchical sampling procedure is efficient with $O(K(\log L + \log K))$ complexity.
5. **Clear Complexity Improvements:** Table I provides a clear summary of improvements: time dependence improves from $O(t^{\{1+1/K\}})$ to $O(t^{\{1+1/(2K+1)\}})$ for PTSC, while NCC achieves $O(n^{\{1+2/(2K+1)\}})$ system-size scaling with $O(\{-1/(2K+1)\})$ accuracy dependence.

Weaknesses

1. **Incomplete Treatment of Random Sampling Overhead:** The overhead factor mentioned in Proposition 1 is significant but receives insufficient analysis. For practical applications, the paper should clarify when the sampling overhead outweighs the gate count improvements. The statement “To make the algorithm efficient, we need to set α to be a constant” (p. 2) obscures the actual trade-off space. The comparison in Fig. 8 sets $\alpha = 2$, implying a factor of 16 overhead that should be explicitly incorporated into the resource comparisons. [FIELD-NORM UNVERIFIED: While variance overhead is a known consideration in randomized quantum algorithms (see Berry et al. 2015, PRL 114), I could not verify a specific threshold or community standard for when this overhead is considered acceptable vs. prohibitive.]
2. **Missing Comparison with Quantum Signal Processing:** The paper positions itself against Taylor-series LCU methods (Refs. 25-27) but does not adequately address quantum signal processing (QSP) approaches, particularly the QSVT framework (Gilyén et al. 2019). Since QSP achieves optimal query complexity with different circuit structures, a theoretical or numerical comparison would strengthen the contribution claims. The brief mention in Appendix G is insufficient for a paper claiming state-of-the-art improvements. Reference needed: Gilyén, Su, Low, Wiebe, “Quantum singular value transformation and beyond,” STOC 2019.
3. **Limited Empirical Validation:** All numerical results (Fig. 8) compare analytical gate count bounds rather than actual circuit implementations or runtime measurements. For claims of “2 orders of magnitude” improvement, empirical validation on specific Hamiltonians (e.g., Fermi-Hubbard, electronic structure) would be more convincing. The reliance on analytical bounds from Ref. [12] for the Trotter baseline may overestimate the actual improvement since tighter empirical bounds often exist.
4. **Ambiguous Treatment of Coherent vs. Random Implementation:** The paper “primarily focus[es] on the random-sampling implementation” (p. 2) but discusses coherent implementation in Appendix H. The relationship between these two regimes and when each is preferable is unclear. For fault-tolerant applications, coherent implementations are often preferred despite higher ancilla costs.
5. **Nested Commutator Coefficient Estimation:** The padding technique (Eqs. 17-19, Fig. 6) introduces virtual qubits to ensure uniform sampling of nested commutators. While clever, this may introduce practical overhead that is not fully accounted for in the complexity analysis. The assumption that all padded commutators have identical 1-norm (2§) ignores potential structure that could be exploited.

Detailed Comments

Literature Review

- **Coverage:** The paper covers classical Trotter references (Suzuki, Childs) and LCU fundamentals (Childs-Wiebe, Berry et al.) adequately. However, several relevant works are missing:
 - Randomized product formula literature: Campbell (2019, PRL) on random compiling, Childs et al. (2019) on theory of Trotter error with randomization
 - QSP/QSVT framework comparison: Gilyén et al. (2019), Martyn et al. (2021, PRX Quantum)
 - Recent hybrid approaches: Low (2019) on combining techniques
- **Integration quality:** The literature is well-integrated into the narrative, with clear positioning against existing methods. The progression from Trotter to post-Trotter to the proposed hybrid is logical.
- **Research gap argument:** The gap identification—Trotter has poor accuracy scaling while LCU has poor system-size scaling—is well-established and the proposed solution directly addresses it.

Theoretical Framework

- **Appropriateness:** The theoretical framework combining product formulas with LCU is appropriate and well-constructed. The use of BCH expansion for error analysis and Euler’s formula for 1-norm reduction are technically sound.
- **Application depth:** The framework is applied deeply, with detailed analysis of order conditions (Lemma 1, Proposition 5), 1-norm bounds (Propositions 3, 4, 6), and complexity proofs (Theorems 1, 2).
- **Alternative frameworks:** The paper could discuss whether similar improvements could be achieved through other hybrid approaches, such as combining Trotter with quantum walks or QSP.

Academic Argument Quality

- **Factual accuracy:** The technical claims appear accurate. The BCH expansion, Trotter order conditions, and LCU framework are correctly presented.
- **Argument logic:** The logical progression is sound: (1) Trotter error has specific structure, (2) leading-order terms are anti-Hermitian, (3) Euler’s formula can reduce 1-norm, (4) this leads to improved complexity. Minor gap: the jump from 1-norm bounds to gate complexity could be more explicit.
- **Terminology precision:** Terms are used consistently with field conventions. “Commutator scaling” follows Childs et al. (2021), “1-norm” follows LCU literature conventions.

Contribution to the Field

- **Incremental contribution:** The contribution is genuine and non-trivial. The specific improvements are:
 - Time scaling: $O(t^{\{1+1/K\}})$ $O(t^{\{1+1/(2K+1)\}})$
 - Accuracy: polynomial logarithmic for PTSC
 - System size preservation for NCC
 - Single-ancilla implementation
- **Positioning:** The paper correctly positions itself as bridging Trotter and LCU paradigms rather than superseding either.
- **Overclaiming risk:** The “2 orders of magnitude” claim (p. 1) refers to analytical bounds and should be qualified. The comparison baseline (4th-order Trotter with analytical bounds from Ref. [12]) is conservative; tighter empirical bounds may reduce the claimed improvement.

Missing Key References

- Gilyén, Su, Low, Wiebe, “Quantum singular value transformation and beyond: exponential improvements for quantum matrix arithmetics,” STOC 2019 — essential for positioning against QSP
- Campbell, “Random compiler for fast Hamiltonian simulation,” PRL 123, 070503 (2019) — relevant randomized approach
- Childs, Ostrander, Su, “Faster quantum simulation by randomization,” Quantum 3, 182 (2019) — randomized Trotter analysis

- Wan, Berta, Campbell, “Randomized semi-quantum matrix processing,” arXiv:2307.00226 — recent hybrid approach
- Low, “Hamiltonian simulation with nearly optimal dependence on spectral norm,” STOC 2019 — optimal Hamiltonian simulation bounds

Questions for Authors

1. **Sampling overhead trade-off:** For what parameter regimes (t, n, ϵ) does the $\epsilon = 16$ sampling overhead outweigh the gate count improvements? Can you provide a phase diagram showing when PTSC/NCC is preferable to deterministic methods?
2. **Coherent implementation performance:** For the coherent implementation discussed in Appendix H, what is the expected ancilla overhead and how does the total resource count compare to QSP-based methods?
3. **Empirical validation:** Have you implemented the algorithms on specific Hamiltonians (e.g., Heisenberg, Fermi-Hubbard) to verify the analytical bounds? What is the observed constant factor overhead?
4. **Extension to non-lattice Hamiltonians:** The NCC algorithm assumes locality (lattice structure). How does performance degrade for molecular Hamiltonians with non-local terms?
5. **Higher-order NCC:** The paper presents second-order NCC results in detail but states “we leave precise higher-order NCC gate count analysis for future study” (p. 10). Is there a fundamental obstacle to extending the analysis?

Minor Issues

- Page 1, line 2: “Department of Computer Science” is duplicated in affiliation 4
- Equation numbering is inconsistent in some appendices
- Figure 8 caption could specify the exact Hamiltonian parameters used
- The term “post-Trotter” (p. 1) is informal; consider “beyond-product-formula methods”
- Reference [20] appears very frequently; consider if all citations are necessary or if some could be consolidated
- Algorithm 1 (p. 8) is specific to Heisenberg model; clarify its generalization to other lattice models

4.5 ARS — Phase 1d: Perspective reviewer

Model: claude-opus-4-5-20251101 · Peer Reviewer 3 cross-disciplinary connections, practical impact.

4.5.1 Perspective Review Report (Peer Reviewer 3)

Reviewer Identity I am a researcher in quantum computing systems engineering and practical quantum algorithm implementation, with expertise in bridging theoretical algorithm design and near-term hardware constraints. My perspective comes from the practical implementation side of quantum computing—understanding what makes algorithms deployable on actual quantum devices versus theoretically elegant. I evaluate this paper through the lens of experimental realizability, resource estimation accuracy, and the gap between asymptotic complexity claims and constant-factor practical performance.

Overall Recommendation Accept

Confidence Score 3

Summary Assessment This paper presents Trotter-LCU algorithms that combine product formulas with linear combination of unitaries to achieve improved Hamiltonian simulation performance. From a practical implementation perspective, the work addresses a genuine need: bridging the gap between high-precision post-Trotter methods that require complex coherent implementations and simpler Trotter methods with poor accuracy scaling. The random-sampling implementation using only one ancillary qubit is particularly valuable for near-term and early fault-tolerant devices. However, the paper’s practical claims require careful scrutiny. The gate count comparisons in Figure 8, while showing 2-4 orders of magnitude improvements, rely on analytical bounds rather than empirical circuit simulations. The “easy to implement” framing may understate challenges in classical preprocessing, measurement overhead, and the practical implications of the \mathfrak{t} sampling overhead. The paper would benefit from more explicit discussion of the practical regimes where these algorithms would outperform alternatives—particularly considering that the overhead factors hidden in asymptotic notation matter enormously for near-term applications. Despite these concerns, the core contribution of achieving commutator scaling with improved accuracy dependence through a simple random-sampling scheme represents genuine practical value.

Strengths (3-5 items)

1. **Minimal Quantum Resource Requirements:** The algorithm requires only one ancillary qubit for random-sampling implementation, directly addressing a key constraint for near-term quantum devices. This is a significant practical advantage over coherent LCU implementations that require $O(\log)$ ancilla qubits for amplitude encoding.
2. **Practical Circuit Structure:** The modification to existing Trotter circuits is additive—appending controlled-Pauli or controlled-Pauli-rotation gates after each Trotter segment. This means practitioners can leverage existing optimized Trotter circuit compilers without fundamental restructuring.
3. **Efficient Classical Sampling Algorithms:** The hierarchical sampling procedure (Figures 5, 7, and Algorithm 1) achieves $O(K(\log L + \log K))$ time complexity per sample, which is practically efficient for lattice Hamiltonians. The explicit algorithmic descriptions enable direct implementation.
4. **Commutator Scaling Preservation:** The NCC algorithm retains the favorable system-size scaling of Trotter methods (near-linear in n for lattice systems) while improving accuracy dependence. This addresses a practical limitation of generic post-Trotter methods that scale as $O(n\check{s})$.

Weaknesses (3-5 items)

1. **Sampling Overhead May Dominate in Practice:** The \mathfrak{t} factor in sample complexity (Proposition 1) is described as acceptable when \mathfrak{t} is “constant,” but the paper sets $\mathfrak{t} = 2$ for numerical comparisons, yielding a factor of 16 overhead. For fair comparison with Trotter methods, this overhead should be explicitly quantified in all gate count comparisons. The statement that this “provides an unbiased estimation” of observables glosses over the variance increase that directly impacts practical shot counts.
2. **Classical Preprocessing Costs Unaddressed:** The nested-commutator compensation algorithm requires computing commutator norms (Cs) and implementing the “padding” procedure for inhomogeneous Hamiltonians. For Hamiltonians beyond the simple Heisenberg model, these classical preprocessing costs could be substantial. The paper does not bound or discuss these costs, which is problematic for practitioners evaluating total resource requirements.
3. **Gate Count Comparisons Use Analytical Bounds:** Figure 8 compares against “the best analytical bound of fourth-order Trotter formula” rather than empirically optimized circuits. Randomized compilation techniques, adaptive Trotter ordering, and other practical optimizations can significantly reduce Trotter circuit depth beyond analytical worst-case bounds. The claimed “2 orders of magnitude” improvement may be optimistic.
4. **Regime of Practical Advantage Unclear:** The paper does not clearly delineate when these algorithms would be preferred over alternatives. For short times, standard Trotter suffices. For very high precision, coherent LCU/QSP may be better despite hardware complexity. The “sweet spot” where random-sampling Trotter-LCU provides practical advantage needs explicit characterization in terms of $(\mathfrak{t}, , n, L)$ parameter regimes.

Detailed Comments

Assumption Audit

- **Explicit assumptions:** The paper assumes single-qubit Z-rotation gates ($R_z()$) are the dominant cost on fault-tolerant quantum computers, which is standard but masks significant variation—magic state distillation costs depend heavily on the rotation angle precision, not just gate count. The assumption that Hamiltonians decompose into efficiently implementable two-local terms (Eq. 10) is stated but may not hold for all practical applications (e.g., molecular Hamiltonians with many-body interactions).
- **Implicit assumptions:** The paper implicitly assumes that the classical random sampling can be performed efficiently in parallel with quantum circuit execution (“we can perform the classical sampling during the quantum circuit implementation or even generate the sampled Pauli matrices before the implementation”). This requires tight classical-quantum integration that may not be available on all platforms. Additionally, the comparison framework assumes observable estimation as the end goal—for applications requiring the full quantum state (e.g., as input to subsequent algorithms), the random-sampling approach is inappropriate.
- **Paradigmatic assumptions:** The paper operates within the paradigm that asymptotic complexity improvements translate to practical advantage. This paradigm often fails for quantum algorithms where constant factors, compilation overhead, and error correction costs dominate. The “orders of magnitude” claims would be more compelling if validated against concrete problem instances with explicit resource accounting.

Cross-Disciplinary Connections

- **Parallel research:** In quantum error correction and fault-tolerant compilation, similar “compilation by randomization” ideas appear in randomized compiling and probabilistic error cancellation. The relationship between this work and these error mitigation techniques is not discussed but could be valuable—particularly whether Trotter-LCU has favorable noise resilience properties.
- **Borrowing opportunities:** From the perspective of optimization algorithms, this work could benefit from adaptive methods that dynamically adjust the Trotter order K or truncation order sc based on intermediate measurement results. The fixed-order approach may leave performance on the table compared to adaptive strategies.
- **Methodological borrowing:** Monte Carlo variance reduction techniques from computational statistics (importance sampling, control variates, stratified sampling) could potentially reduce the \dagger sampling overhead. The paper’s random sampling approach is basic; sophisticated sampling strategies might improve practical performance.

Practical Impact

- **Real-world application:** For early fault-tolerant quantum computers with limited ancilla qubits, this algorithm provides a concrete path to improved Hamiltonian simulation accuracy. The explicit sampling algorithms (Algorithm 1, Figures 5, 7) are implementation-ready.
- **Implementation feasibility:** The main barriers are: (1) the “padding” procedure for NCC requires Hamiltonian-specific preprocessing; (2) the Hadamard-test-type circuit requires mid-circuit measurement and reset (Figure 10c), which has hardware-dependent fidelity; (3) the \dagger sampling overhead means practical experiments require $16E$ more shots than standard Trotter at minimum.
- **Stakeholders:** The paper primarily addresses algorithm designers and quantum software developers. Hardware considerations (decoherence during the extended circuit, measurement/reset fidelity) are not discussed. Experimentalists implementing these algorithms would need guidance on error sensitivity.

Broader Implications

- **Ethical dimensions:** No significant ethical concerns. The work is fundamental algorithm research.
- **Social impact:** As with all quantum simulation algorithms, success would accelerate materials science, drug discovery, and other applications. No differential social impact concerns.
- **Future directions:** From an implementation perspective, the most valuable follow-up would be: (1) empirical validation on actual or simulated quantum hardware with noise; (2) integration with quantum error correction to understand logical error rates; (3) comparison against other random-sampling algorithms (qDRIFT, etc.) on specific benchmark Hamiltonians with full resource accounting.

Cross-Disciplinary Reading Recommendations

- **Campbell, E. (2019).** “Random Compiler for Fast Hamiltonian Simulation.” *Physical Review Letters* **123**, 070503. — The qDRIFT algorithm takes a different random-sampling approach; comparison would clarify the relative advantages of Trotter-LCU versus pure randomized methods.
- **Koczor, B. (2021).** “Exponential Error Suppression for Near-Term Quantum Devices.” *Physical Review X* **11**, 031057. — Discusses variance reduction for expectation value estimation via randomization; techniques may apply to reduce the \dagger overhead.
- **van den Berg, E., et al. (2023).** “Probabilistic error cancellation with sparse Pauli-Lindblad models on noisy quantum processors.” *Nature Physics* **19**, 1116. — Demonstrates practical random-sampling quantum algorithms on real hardware; provides context for implementation challenges.
- **Childs, A. M., & Su, Y. (2019).** “Nearly optimal lattice simulation by product formulas.” *Physical Review Letters* **123**, 050503. — The tight Trotter bounds for lattice systems that this paper builds upon; important for understanding where the baseline comparisons come from.

Questions for Authors

1. For the numerical comparisons in Figure 8, have you considered comparing against empirically optimized Trotter circuits (e.g., with randomized ordering or recompilation) rather than analytical worst-case bounds? The claimed improvement magnitude may be sensitive to this choice.
2. What is the practical variance of the observable estimator as a function of \dagger ? The \dagger scaling suggests significant shot overhead; explicit numerical examples showing the required shot count for practical precision (e.g., 1% relative error on energy) would help practitioners evaluate the algorithm.
3. For the NCC algorithm on general (non-lattice) Hamiltonians, what is the classical preprocessing complexity to compute the commutator norms and implement the padding procedure? Is this tractable for, e.g., molecular Hamiltonians with hundreds of qubits?
4. How do these algorithms compare to qDRIFT and related randomized product formula methods in the same (n, t, \dagger) parameter regimes? A direct comparison would help position this work within the broader landscape of randomized Hamiltonian simulation.

Minor Issues

- The notation $\text{Pol}(s; x)$ in Eq. (4) should be $\text{Poi}(s; x)$ for consistency with later usage.
- The term “Hamiltonian simulation” in the abstract could clarify whether coherent simulation or observable estimation is the primary goal, given the random-sampling focus.
- The Lambert W function reference (Lemma 7, Appendix F) is somewhat technical for the main text discussion; a simpler bound approximation might improve accessibility.
- Figure 6 caption refers to “virtual ancilla qubits” but the physical interpretation of “padding” could be clearer for implementers.

4.6 ARS — Phase 1e: Devil’s Advocate

Model: claude-haiku-4-5-20251001 · Core argument challenges, logical fallacy detection, counter-arguments.

4.6.1 Devil’s Advocate Review

Strongest Counter-Argument This paper presents a clever hybrid of Trotter formulas and LCU compensation that claims to capture “the best of both worlds.” However, the central value proposition is fundamentally undermined by a practical reality the authors systematically downplay: **the t sampling overhead inherent to their random-sampling LCU implementation may completely erase the theoretical gate-count advantages in any realistic experimental setting.**

The authors set $t = 2$ for their numerical comparisons (Section II.E), leading to a $16\mathbb{E}$ multiplicative overhead in sample complexity compared to deterministic implementations. Yet their gate-count comparisons (Figure 8) are presented as single-shot counts, obscuring this overhead. A fair comparison would multiply PTSC/NCC gate counts by 16 (or more generally t) when contrasting against coherent Trotter implementations. Under this adjustment, the “2 orders of magnitude improvement” over fourth-order Trotter (claimed in the abstract) reduces to approximately $6\mathbb{E}$ improvement at best—and this advantage may flip entirely for moderate precision requirements where Trotter’s deterministic nature becomes preferable.

Furthermore, the claimed “easy implementation” advantage is predicated on avoiding multi-controlled Toffoli gates, but the algorithm introduces controlled-Pauli rotations e^{iP} at each compensation step. On near-term devices with limited connectivity and high rotation-gate errors, these controlled rotations—especially with Pauli strings of weight up to $O(s_c)$ —may be operationally harder than the Toffoli gates they supposedly replace. The authors provide no experimental validation, gate-error analysis, or hardware-noise simulation to support their implementability claims.

The “commutator scaling” advantage for NCC on lattice Hamiltonians is real but narrow: it applies only to systems with strong locality structure (nearest-neighbor interactions). For the chemistry Hamiltonians mentioned as a target application (Section II.B), the $O(n^3)$ dependence that plagues post-Trotter methods would equally afflict NCC, since molecular Hamiltonians lack the geometric locality that enables commutator savings. The paper’s scope claims exceed its demonstrated results.

Issue List

CRITICAL

#	Dimension	Issue Description	Location	Field-Norm Boundary	Evidence-Crossing Rationale
C1	Core Thesis Challenge	The t sampling overhead (Proposition 1) is not accounted for in gate-count comparisons. Setting $t=2$ implies $16\mathbb{E}$ more circuit executions than deterministic Trotter. Figure 8’s “gate count” comparisons are misleading because they compare single-shot costs without normalizing for total experimental resources.	Section II.E, Figure 8, Proposition 1		

#	Dimension	Issue Description	Location	Field-Norm Boundary	Evidence-Crossing Rationale
C2	Logic Chain Validation	The claim that PTSC/NCC are “easy to implement” (Abstract, Table I) rests on avoiding multi-controlled Toffoli gates. But controlled-Pauli rotations on weight- $O(s_c)$ Pauli strings require $O(s_c)$ CNOT gates per rotation (Figure 11b). For $s_c = O(\log(1)/\log \log(1/))$, this creates a non-trivial circuit depth that is never quantified against Toffoli-gate costs in post-Trotter methods. The “easy implementation” claim is asserted without comparative evidence.	Abstract, Section II.E, Figure 11, Table I		

MAJOR

#	Dimension	Issue Description	Location	Field-Norm Boundary	Evidence-Crossing Rationale
M1	Overgeneralization Check	The paper claims applicability to “quantum chemistry Hamiltonians with large L ” (Section II.B), but all analytical bounds and numerical examples are for lattice models with geometric locality. Molecular Hamiltonians lack this locality structure, and the NCC commutator-scaling advantage would not apply. No chemistry-relevant numerical demonstration is provided.	Section II.B, Section VI		
M2	Alternative Paths Analysis	The paper dismisses coherent LCU implementation in favor of random sampling, stating it is “more suitable to implement in near-term devices” (Section III.B). However, recent work on early fault-tolerant quantum computing shows that coherent implementations with moderate ancilla counts may be preferable to high-variance random-sampling schemes. The authors do not compare variance-normalized costs.	Section III.B		

#	Dimension	Issue Description	Location	Field-Norm Boundary	Evidence-Crossing Rationale
M3	Cherry-Picking Detection	The Trotter error comparison baseline uses the analytical bound from Ref. [12], which is known to be loose by orders of magnitude for structured Hamiltonians. Tighter commutator-aware Trotter bounds (Ref. [20], which the authors cite for NCC) would reduce the apparent advantage of PTSC. The comparison selectively uses the weakest Trotter baseline to maximize apparent improvement.	Section II.E, Figure 8(a,b)		
M4	Logic Chain Validation	Proposition 4’s bound $\hat{t}(p)(x) \exp((e+2/9)(2x))$ contains an unexplained constant $(e+2/9 \cdot 2.94)$. The derivation in Eq. (80) involves multiple inequalities but does not justify why this particular constant is tight or whether it hides significant slack that would affect practical performance.	Proposition 4, Eq. (80)		
M5	“So What?” Test	The paper’s primary contribution—improved time scaling from $1+1/K$ to $1+1/(2K+1)$ —yields modest practical improvements. For $K=4$ (fourth-order Trotter), time scaling improves from $\hat{t}^{1.25}$ to $\hat{t}^{1.11}$. For $t=100$, this is a factor of ~ 1.6 improvement in gate count, which may be overwhelmed by the $\hat{t}=16$ sampling overhead. The incremental contribution is smaller than headline claims suggest.	Section IV.C, Theorem 1		

MINOR

#	Dimension	Issue Description	Location
m1	Logic Chain Validation	The notation “Department of Computer Science, Department of Computer Science” in the affiliation list appears to be a typographical error.	Author affiliations
m2	Confirmation Bias Detection	All numerical comparisons show PTSC/NCC outperforming baselines. No parameter regime is identified where Trotter methods would be preferred, creating an impression of universal superiority that is inconsistent with the \hat{t} overhead reality.	Figure 8, Section II.E

#	Dimension	Issue Description	Location
m3	Stakeholder Blind Spots	The paper does not discuss hardware constraints of specific near-term platforms (e.g., superconducting qubits, trapped ions). “Easy implementation” is platform-agnostic but actual difficulty is platform-specific.	Throughout

Ignored Alternative Explanations/Paths

1. **Hybrid coherent/random implementation:** The authors present coherent and random-sampling LCU as mutually exclusive (Section III.B), but a hybrid approach using coherent implementation for low-order terms and random sampling only for residuals could reduce variance while maintaining simplicity. This is not explored.
2. **Randomized compiling baselines:** Recent work on randomized compilation of Trotter circuits achieves similar error-suppression effects through Pauli twirling. These methods are simpler than LCU compensation and may achieve comparable accuracy improvements for similar sampling costs. No comparison is provided.
3. **Tensor network / classical simulation baselines:** For the lattice Hamiltonians with $n=12-50$ qubits (Figure 8c), tensor network methods may be competitive for the claimed precision levels. The “quantum advantage” context is unstated.

Missing Stakeholder Perspectives

- **Experimentalists:** No discussion of how controlled-Pauli rotations would be implemented on specific hardware, what gate fidelities are required, or how mid-circuit measurement (Figure 2c) affects coherence.
- **Fault-tolerance architects:** The paper claims suitability for “near-term or early fault-tolerant quantum computer” (Section I) but does not address magic-state distillation costs for $Rz()$ gates in fault-tolerant settings.

Observations (Non-Defects)

- The nested-commutator structure (Section V) is a genuine theoretical contribution that connects Trotter error analysis to LCU compensation in a novel way.
- The “Hamiltonian padding” technique (Section II.D, Eq. 17-19) to ensure uniform sampling is a practical insight that could benefit other randomized algorithms.
- Algorithm 1 provides an unusually concrete sampling procedure that enhances reproducibility relative to typical theory papers.

4.7 ARS — Phase 2: Editorial synthesizer

Model: claude-haiku-4-5-20251001 · Aggregates all reviews Editorial Decision + Revision Roadmap.

I’ll synthesize the 5 reviewer reports to produce the Editorial Decision Letter and Revision Roadmap. Let me carefully analyze the reviewer cards and identify consensus, disagreements, and construct the final editorial package.

Editorial Decision Package

4.7.1 Part 1: Editorial Decision Letter

Dear Author(s),

Thank you for submitting your manuscript titled “Simple and high-precision Hamiltonian simulation by compensating Trotter error with linear combination of unitary operations” to PRX Quantum. Your manuscript has been reviewed by the Editor-in-Chief and three independent peer reviewers with expertise in mathematical rigor, quantum simulation algorithms, and practical implementation.

Decision: Minor Revision

Consensus Analysis

Points of Agreement (Consensus)

[CONSENSUS-4] Issues (All 4 reviewers agree)

- **SC-1: t sampling overhead is inadequately addressed in gate-count comparisons.** All reviewers note that the $t = 16$ overhead (with $t = 2$) from random sampling is not explicitly incorporated into the Figure 8 comparisons, potentially overstating practical advantages. (EIC \checkmark Weaknesses-2; R1 \checkmark D4 significance; R2 \checkmark Weaknesses-1; R3 \checkmark Weaknesses-1; DA \checkmark C1)
- **SC-2: Numerical comparisons use analytical bounds rather than empirical validation.** All reviewers observe that Figure 8 comparisons pit the proposed algorithms against analytical Trotter bounds from Refs. [12, 20], which may overstate improvements since empirical Trotter performance can be significantly better. (EIC \checkmark Weaknesses-1; R2 \checkmark Weaknesses-3; R3 \checkmark Weaknesses-3; DA \checkmark M3)
- **SC-3: The core algorithmic contribution—combining Trotter with LCU compensation using order-pairing—is novel and sound.** All reviewers acknowledge the conceptual elegance and technical validity of the approach. (EIC \checkmark Strengths-1; R1 \checkmark NOVELTY evaluation; R2 \checkmark Strengths-1,2; R3 \checkmark Strengths-2)

[CONSENSUS-3] Issues (3 of 4 reviewers agree)

- **SC-4: Missing comparison with QSP/QSVT methods.** R2, R3, and DA note the absence of comparison with quantum signal processing approaches (Gilyén et al. 2019). EIC did not raise this explicitly. (R2 \checkmark Weaknesses-2; R3 \checkmark Cross-Disciplinary; DA \checkmark M2)
- **SC-5: Higher-order NCC analysis is incomplete.** EIC, R2, and R3 note that only second-order NCC results are presented in detail, with higher-order analysis deferred to future work. R1 did not specifically flag this. (EIC \checkmark Weaknesses-3; R2 Question 5; R3 implicit in practical regime discussion)
- **SC-6: Paper length and density could be reduced.** EIC, R2 (implicit), and R3 suggest streamlining the main text. R1 did not raise this as a concern. (EIC \checkmark Weaknesses-4)

Points of Disagreement

Disagreement 1: Severity of the t overhead concern

- **R3 (Perspective) and DA:** View this as potentially undermining the paper’s central claims; DA elevates it to CRITICAL status.
- **EIC and R2 (Domain):** Acknowledge it as a limitation requiring clarification but not as invalidating the contribution.
- **R1 (Methodology):** Notes it as a significance concern but defers to domain expertise.

Editor’s Resolution: The t overhead is a legitimate concern that requires explicit quantification, but it does not invalidate the theoretical contribution. **Required action:** Authors must add a dedicated analysis clarifying the crossover point where random-sampling advantages outweigh the overhead, and explicitly note the t factor when presenting gate-count comparisons. This is a clarification/supplement requirement, not grounds for rejection.

Rationale: The theoretical scaling improvements (time dependence from $t^{(1+1/K)}$ to $t^{(1+1/(2K+1))}$) are mathematically established. The random-sampling implementation is one implementation pathway; the paper’s value does not depend solely on it dominating in all regimes. The request is for transparency, not for proving universal superiority.

Disagreement 2: “Easy implementation” claims

- **DA:** Argues that controlled-Pauli rotations on weight- $O(s_c)$ Pauli strings may not be easier than Toffoli gates (CRITICAL issue C2).
- **R3:** Notes implementation feasibility concerns but considers the single-ancilla requirement a genuine advantage (Strength 1).
- **R2 and EIC:** Accept the “easy implementation” framing as reasonable given the avoided multi-controlled gates.

Editor’s Resolution: The “easy implementation” characterization is contextually valid for the single-ancilla random-sampling approach but requires qualification. **Required action:** Authors should clarify that “easy implementation” refers specifically to the ancilla-efficient random-sampling variant and acknowledge that controlled-Pauli rotation synthesis has its own complexity ($O(\text{wt}(P))$ CNOTs per rotation). This is not a retraction of the claim but a precision improvement.

Rationale: R3’s Confidence Score of 3 (moderate—outside primary expertise) on implementation details, combined with EIC and R2’s acceptance (Confidence 4), suggests the claim is defensible but would benefit from precision.

Disagreement 3: Applicability to chemistry Hamiltonians

- **DA:** Claims the paper overgeneralizes by mentioning chemistry applications (M1) when all examples are lattice models.
- **R2:** Notes extension to non-lattice Hamiltonians is an open question (Question 4) but does not flag this as a major flaw.
- **EIC and R3:** Do not raise this as a significant concern.

Editor’s Resolution: The paper’s mention of chemistry applications (Section II.B) is aspirational rather than demonstrated. **Suggested action:** Authors should either provide a brief chemistry Hamiltonian example or qualify the chemistry claims as future work direction. This is a P2 (should fix) item, not required.

Rationale: DA’s concern has merit but the paper’s primary contribution is clearly demonstrated for lattice systems. The chemistry mention is a brief aside, not a central claim.

Decision Rationale This manuscript presents a genuinely novel contribution to Hamiltonian simulation by demonstrating that Trotter formulas and LCU methods can be synergistically combined rather than treated as competing approaches. The core theoretical results—improved time scaling from $1+1/K$ to $1+1/(2K+1)$ and logarithmic accuracy dependence for PTSC—are mathematically sound and represent a meaningful advance over existing methods.

The reviewers unanimously acknowledge the technical merit of the “order-pairing” technique using Euler’s formula and the preservation of commutator scaling in the NCC algorithm. The practical value of a single-ancilla implementation pathway is recognized, though the random-sampling overhead requires more transparent treatment.

The primary concerns are presentational and comparative rather than foundational: 1. Gate-count comparisons need to explicitly account for sampling overhead 2. Baselines should acknowledge that analytical Trotter bounds are conservative 3. Higher-order NCC deserves at least rough estimates 4. QSP/QSVT comparison would strengthen positioning

None of these concerns challenge the validity of the theoretical contribution or its potential practical value in appropriate regimes. The paper should be accepted following minor revisions that improve transparency and contextualization.

Summary of Key Issues

1. **‡ sampling overhead transparency** — All reviewers, CONSENSUS-4
 2. **Analytical bounds vs. empirical validation** — All reviewers, CONSENSUS-4
 3. **Missing QSP/QSVT comparison** — R2, R3, DA, CONSENSUS-3
 4. **Incomplete higher-order NCC analysis** — EIC, R2, R3, CONSENSUS-3
 5. **“Easy implementation” claim precision** — DA (critical), R3 (notes concerns)
-

4.7.2 Part 2: Revision Roadmap

The Sub-Claim(s) column carries the Step 1b sub_claim_id(s) each item traces to.

Required Revisions (Must Fix)

#	Revision Item	Sub-Claim(s)	Source	Pri- ority	Estimated Effort
R1	Add explicit discussion of t sampling overhead in Section II.E and Figure 8 caption. Quantify total resource cost (gates \otimes samples) alongside single-shot gate counts.	SC-1	EIC, R1, R2, R3, DA	P1	2-3 days
R2	Add subsection or paragraph clarifying when random-sampling PTSC/NCC outperforms deterministic Trotter, including crossover analysis in terms of $(t, n,)$ parameters.	SC-1	EIC Question 2, R2 Question 1, R3 Question 2	P1	3-4 days
R3	Acknowledge in Section II.E that Figure 8 comparisons use analytical Trotter bounds and note that empirical performance may reduce the apparent improvement.	SC-2	EIC, R2, R3, DA	P1	1 day
R4	Fix duplicate “Department of Computer Science” in affiliation 4.	—	EIC, DA	P1	<1 hour

Suggested Revisions (Should Fix)

#	Revision Item	Sub-Claim(s)	Source	Prior- ity	Estimated Effort
S1	Add brief comparison or discussion of QSP/QSVT methods, explaining why Trotter-LCU offers complementary advantages (e.g., single ancilla, commutator scaling). Add Gilyén et al. 2019 reference.	SC-4	R2, R3, DA	P2	2-3 days
S2	Provide rough estimates or bounds for fourth-order NCC performance, even if detailed analysis is deferred.	SC-5	EIC, R2	P2	2-3 days
S3	Qualify “easy implementation” claims (Abstract, Table I) to specify this refers to the random-sampling variant with single ancilla. Note the $O(\text{wt}(P))$ CNOT complexity for controlled-Pauli rotations.	—	DA, R3	P2	1 day
S4	Clarify Equation (21) coefficient of 4 in Heisenberg Hamiltonian normalization.	—	EIC	P2	<1 hour
S5	Streamline main text to ~15-18 pages by moving detailed derivation steps to appendices.	SC-6	EIC	P2	3-5 days
S6	Add missing references: Gilyén et al. (STOC 2019), Campbell (PRL 2019), Childs et al. randomized Trotter (Quantum 2019).	—	R2	P2	1 day
S7	Clarify that chemistry Hamiltonian applicability (Section II.B) is aspirational, or provide a brief molecular Hamiltonian example.	—	DA, R2	P2	1-2 days

#	Revision Item	Sub-Claim(s)	Source	Prior-ity	Estimated Effort
S8	Distinguish notation for LCU accuracy vs. simulation accuracy , or clarify usage.	—	EIC	P2	<1 hour

Revision Checklist (Checkable List)

Priority 1 — Structural Revisions (Estimated total effort: 6-8 days)

- R1: Explicitly incorporate $t = 16$ overhead in all gate-count discussions and Figure 8 comparisons
- R2: Add crossover analysis for when random-sampling approach is preferable
- R3: Caveat that Figure 8 uses analytical bounds; empirical Trotter may perform better
- R4: Correct duplicate affiliation text

Priority 2 — Content Supplementation (Estimated total effort: 10-14 days)

- S1: Add QSP/QSVT comparison discussion and references
- S2: Provide rough higher-order NCC estimates
- S3: Qualify “easy implementation” with precision about circuit complexity
- S4: Clarify Heisenberg Hamiltonian normalization
- S5: Condense main text, move detailed derivations to appendices
- S6: Add missing key references
- S7: Qualify or demonstrate chemistry Hamiltonian claims
- S8: Clarify notation usage

Priority 3 — Text and Formatting (Estimated total effort: 2-3 days)

- Fix typo “allowss” “allows” (p. 16)
- Correct Poi vs. Pol notation inconsistency (Eq. 4)
- Improve Figure 8 legend clarity re: which bounds are used
- Ensure equation numbering consistency in appendices
- Consolidate excessive citations to Ref. [20] where appropriate

Revision Deadline Recommended: 4-6 weeks (Minor revision timeline)

Response Letter Template Authors should use the journal’s standard response format to address each numbered revision item, explaining how each concern was addressed or providing justification for alternative approaches.

4.7.3 Part 3: Reviewer Report Summary (Appendix)

EIC Report Summary

- **Recommendation:** Minor Revision | **Confidence:** 4
- **Key Point:** Elegant conceptual contribution combining Trotter and LCU; practical claims need grounding through explicit sampling overhead analysis.

Reviewer 1 (Methodology) Summary

- **Recommendation:** (Pass on correctness, warn on significance without overhead clarity) | **Confidence:** 4
- **Key Point:** Mathematical framework is sound; 1-norm bounds and proofs are rigorous; significance claims depend on fair comparison methodology.

Reviewer 2 (Domain) Summary

- **Recommendation:** Minor Revision | **Confidence:** 4
- **Key Point:** Genuine algorithmic advance bridging Trotter and LCU paradigms; needs QSP comparison and explicit treatment of sampling overhead trade-offs.

Reviewer 3 (Perspective) Summary

- **Recommendation:** Accept | **Confidence:** 3
- **Key Point:** Practical single-ancilla implementation is valuable; regime of practical advantage needs clearer delineation; variance overhead concerns warrant attention.

Devil's Advocate Summary

- **Critical Issues:** † sampling overhead (C1) and unsubstantiated “easy implementation” claims (C2)
- **Assessment:** DA’s sampling overhead concern is corroborated by all reviewers (CONSENSUS-4). The “easy implementation” challenge is partially addressed by reviewer acceptance of single-ancilla value but warrants qualification.

Editorial Note: The Devil’s Advocate raised the sampling overhead as a CRITICAL issue (C1). This concern is fully corroborated by all four primary reviewers and has been incorporated as Required Revision R1/R2. The DA’s second critical issue (C2, “easy implementation”) receives partial corroboration from R3’s concerns but is viewed by EIC and R2 as requiring qualification rather than retraction. The EIC arbitrates this as a P2 suggested revision (S3) rather than a P1 required revision, given the genuine single-ancilla advantage acknowledged by multiple reviewers.

5 ARC peer_review + quality_gate — full per-stage output

Every subsection below is ARC output (`_run/arc/`). ARC is a 23-stage autonomous research pipeline (topic → empirical paper); for this head-to-head we route only the two paper-review stages so the comparison is review-vs-review.

5.1 ARC — Stage 1: Peer review

Model: `claude-haiku-4-5-20251001` · Simulates 2+ reviewer perspectives (A, B), checks methodology-evidence consistency, flags fabrication. Markdown output.

Peer Review: Simple and high-precision Hamiltonian simulation by compensating Trotter error with LCU

5.1.1 Paper Summary

This paper proposes composite Hamiltonian simulation algorithms that combine Trotter formulas with Linear Combination of Unitaries (LCU) to achieve improved accuracy and time scaling. The key contributions are: (1) Paired Taylor-Series Compensation (PTSC) which exponentially improves accuracy scaling, and (2) Nested-Commutator Compensation (NCC) which maintains near-optimal system-size dependence while quadratically improving accuracy over Trotter methods.

Reviewer A

5.1.2 Expertise

Quantum algorithms, Hamiltonian simulation, fault-tolerant quantum computing

5.1.3 Summary Assessment

This is a well-written theoretical paper that presents a clever combination of Trotter and LCU methods for Hamiltonian simulation. The mathematical framework is rigorous and the results appear technically sound. However, there are concerns about the experimental validation and some gaps between claims and evidence.

5.1.4 Strengths

- Novel algorithmic contribution:** The idea of using LCU to compensate Trotter error rather than replacing Trotter entirely is creative and practically motivated. The “order pairing” technique using Euler’s formula (Eq. 7) to double the order of $t_x - 1$ is an elegant insight.
- Strong theoretical foundations:** The paper provides rigorous proofs for both PTSC (Theorem 1) and NCC (Theorem 2) algorithms with explicit gate complexity bounds. The analysis connecting commutator structure to system-size scaling is thorough.
- Practical implementation considerations:** The hierarchical sampling algorithms (Fig. 5, Fig. 7) and the “padding” technique for non-homogeneous Hamiltonians demonstrate attention to practical implementation details.
- Clear exposition:** The paper does an excellent job building intuition through the progressive examples in Fig. 3 and Fig. 4 before presenting general results. The comparison table (Table I) effectively summarizes the contributions.
- Significant complexity improvements:** The claimed 2-4 orders of magnitude improvement in accuracy for the same gate count (Fig. 8c) over state-of-the-art Trotter methods is substantial if accurate.

5.1.5 Weaknesses

Major Issues

1. **CRITICAL: Lack of experimental/numerical validation evidence**

- The paper presents gate count comparisons in Fig. 8 but **no actual quantum simulation results** are provided.
- The gate counts appear to be analytical estimates based on complexity bounds rather than empirical measurements from implemented circuits.
- **No evidence of actual code execution, numerical simulations, or statistical trials** is presented in the paper to validate the theoretical claims.
- The paper should clarify whether Fig. 8 represents analytical bounds or measured gate counts from implementations.

2. **Missing implementation details for reproducibility:**

- Algorithm 1 provides pseudocode for first-order NCC sampling, but no complete implementation details for higher-order algorithms.
- The paper does not specify how the rotation angles in Pauli rotations are computed in practice with finite precision.
- No discussion of numerical stability issues in the sampling procedures.

3. **Limited baseline comparisons:**

- Comparison is primarily against fourth-order Trotter with analytical bounds from Ref. [12, 20].
- No comparison with other recent post-Trotter methods like truncated Taylor series [25] or QSP [27, 28] in the main text (relegated to Appendix G).
- The $t = 2$ setting for fair comparison may not be optimal for all methods.

4. **Incomplete analysis of practical overheads:**

- The t sampling overhead (Proposition 1) means the method requires $16E$ more samples when $t = 2$. This is mentioned but not adequately discussed in context.
- The paper truncates at “the gate complexity estimation is then converted to” (line cut off), leaving analysis incomplete.

Minor Issues

5. **Clarity issues in notation:** The vector notation (e.g., r , p) is sometimes inconsistent between sections.
6. **Missing error analysis:** While truncation error is bounded (Propositions 3, 4), the accumulation of errors over many segments is not rigorously analyzed beyond the product formula (Proposition 2).

5.1.6 Questions for Authors

1. Have you implemented the algorithms and run numerical experiments? If so, what are the actual gate counts vs. the analytical bounds?
2. How does the performance degrade when using finite-precision rotation angles?
3. Can you provide explicit comparisons with QSP-based methods for the lattice Hamiltonian examples?

5.1.7 Actionable Revisions

1. **Add numerical validation:** Implement the algorithms and provide actual circuit simulations validating the gate count estimates.
2. **Expand baseline comparisons:** Include direct comparisons with QSP and truncated Taylor series in the main text.
3. **Discuss the t overhead more thoroughly:** Provide guidance on when the method is practically advantageous given this sampling cost.
4. **Complete the truncated analysis:** The paper appears cut off in Section IV C.

5.1.8 Recommendation

Score: 6/10 - Weak Accept

The theoretical contributions are solid and the algorithmic ideas are novel. However, the lack of empirical validation beyond analytical gate count estimates is a significant weakness for a methods paper. The paper would be strengthened considerably by numerical experiments demonstrating the practical performance of the algorithms.

Reviewer B

5.1.9 Expertise

Quantum computing theory, complexity analysis, near-term quantum algorithms

5.1.10 Summary Assessment

This paper presents theoretically interesting algorithms combining Trotter and LCU methods. While the mathematical framework is sophisticated, I have concerns about the completeness of the presentation and the gap between theoretical claims and supporting evidence.

5.1.11 Strengths

1. **Addresses a real practical problem:** The tension between Trotter’s implementation simplicity and post-Trotter’s accuracy is a genuine challenge. The proposed hybrid approach is well-motivated.
2. **Comprehensive theoretical analysis:** The paper provides complexity bounds for multiple algorithm variants (0th, Kth-order PTSC; Kth-order NCC) with explicit dependence on all relevant parameters.
3. **Commutator structure exploitation:** The NCC algorithm’s ability to maintain $O(n^{\{1+2/(2K+1)\}})$ scaling for lattice Hamiltonians while improving accuracy is theoretically appealing.
4. **Practical sampling procedures:** The hierarchical sampling avoiding full expansion of nested commutators (Sections II.D) is important for practical implementation.
5. **Good presentation quality:** Figures 3-4 provide excellent visual intuition for the order-pairing technique.

5.1.12 Weaknesses

Critical Issues

1. PAPER APPEARS INCOMPLETE

- The paper text cuts off mid-sentence: “Moreover, the accuracy scaling is exponentially im-” at the end of Section IV.C
- Section V (Nested-Commutator Compensation detailed analysis) is referenced but not provided
- Section VI (Conclusion) is missing
- Multiple appendices (A-H) are referenced but not included
- This makes it impossible to fully evaluate the NCC algorithm claims

2. NO EXPERIMENTAL EVIDENCE PROVIDED

- The paper makes quantitative claims (“2 orders of magnitude smaller,” “3 to 4 orders of magnitude higher accuracy”)
- Fig. 8 shows gate count comparisons but the methodology for generating these plots is unclear
- **No code, no numerical experiments, no statistical analysis** is provided to support the claims
- The “gate counting method for fourth-order Trotter with random permutation is based on the analytical bounds in Ref. [12]” - so this is purely theoretical comparison

3. Missing critical algorithmic details:

- How is the truncation order s_c chosen in practice? The bound in Eq. 85 involves Lambert W function - is this computed exactly?
- The “variant circuit” in Fig. 2(c) with measurement and reset is claimed equivalent to Fig. 2(b) - no proof provided
- Error accumulation across segments needs more rigorous treatment

Methodology-Evidence Consistency Issues

4. Claims without supporting evidence:

- Claim: “2 orders of magnitude smaller than the best analytical bound of fourth-order Trotter formula” (Abstract)
 - Evidence: Fig. 8(a,b) shows this, but methodology is analytical bounds comparison, not empirical measurement
 - Claim: “3 to 4 orders of magnitude higher accuracy with the same gate costs” (Abstract)
 - Evidence: Fig. 8(c) - again, based on analytical bounds from Ref. [20], not empirical tests
5. **Venue appropriateness:** This paper is marked for PRX Quantum, which is appropriate for theoretical quantum computing papers. However, even for a theory venue, numerical validation of complexity bounds is standard practice.

Minor Issues

6. **Notation overload:** The paper uses many subscripts and superscripts (e.g., $F_{\{1,s\}}^{\{nc\}}$, $V_{1^{\{p\}}}$, $R_{\{1,s\}}$) which can be difficult to track.
7. **Limited discussion of constants:** The constants in complexity bounds (e.g., $_K = 2 \mathcal{O} 5^{\{K/2-1\}}$) grow quickly with K; practical implications not discussed.
8. **Missing related work:** No comparison with randomized compiling methods or other error mitigation approaches that could achieve similar goals.

5.1.13 Questions for Authors

1. Can you provide the missing sections (V, VI) and appendices?
2. Have you implemented the algorithms? What are the actual vs. predicted gate counts?
3. For Fig. 8, are these analytical bounds or measured values? What is the exact methodology?
4. How sensitive is the performance to the choice of t ? Is $t = 2$ optimal?

5.1.14 Actionable Revisions

1. **CRITICAL: Complete the paper** - Sections V, VI and appendices are essential for evaluation.
2. **Add numerical experiments:** Even for a theory paper, implementing the sampling algorithms and comparing actual circuit depths would strengthen claims significantly.
3. **Clarify Fig. 8 methodology:** Explicitly state whether values are analytical bounds or empirical measurements.
4. **Provide complexity comparison table with explicit constants:** The asymptotic notation hides potentially large constants.
5. **Discuss practical limitations:** When does the t overhead make the method impractical? What system sizes are needed for advantages to manifest?

5.1.15 Recommendation

Score: 5/10 - Borderline (leaning reject)

The theoretical ideas are interesting and potentially impactful. However, the paper is clearly incomplete (missing sections, truncated text), and there is no experimental validation of the theoretical claims. The gap between the sophisticated mathematical analysis and the lack of any empirical evidence is concerning.

For PRX Quantum, I would expect at least numerical validation of gate counts on representative Hamiltonians. The current submission appears to be a draft that was submitted prematurely.

Conditional recommendation: If the authors can provide (1) complete paper with all sections, (2) numerical validation experiments, and (3) clarified methodology for Fig. 8, this could be a solid contribution suitable for publication.

Summary of Critical Issues Requiring Revision

Issue	Severity	Reviewer
Paper incomplete (missing Sections V, VI, appendices)	CRITICAL	B
No experimental/numerical validation	MAJOR	A, B
Fig. 8 methodology unclear (analytical vs empirical)	MAJOR	A, B
Text cuts off mid-sentence	CRITICAL	B
Missing baseline comparisons with QSP	MODER- ATE	A
\dagger overhead implications not discussed	MODER- ATE	A, B

5.2 ARC — Stage 2: Quality gate

Model: claude-haiku-4-5-20251001 · Final JSON verdict score_1_to_10, strengths, weaknesses, required_actions, verdict. Threshold-based PROCEED/REVISE.

The quality gate scores the manuscript **7/10** against a threshold of 7.0. Accept with minor revisions. The paper presents a novel and technically sound approach combining Trotter and LCU methods for Hamiltonian simulation with significant practical improvements. The theoretical contributions are solid and the numerical demonstrations are convincing. However, the manuscript appears incomplete (truncated text) and needs final polish.

Strengths noted. Novel algorithmic framework that synergistically combines Trotter and LCU methods, inheriting advantages of both. Rigorous mathematical analysis with clear proofs of improved time scaling ($1+1/(2K+1)$ vs $1+1/K$) and accuracy dependence. Practical random-sampling implementation requiring only one ancilla qubit with efficient classical sampling procedures. Comprehensive numerical comparisons demonstrating 2-4 orders of magnitude improvement over state-of-the-art Trotter methods. NCC algorithm maintains near-optimal system-size scaling for lattice Hamiltonians while improving accuracy. Clear presentation with helpful figures illustrating the key algorithmic ideas.

Weaknesses noted. The manuscript text is truncated/incomplete - the proof of Theorem 1 cuts off mid-sentence. The \dagger sampling overhead factor compared to standard algorithms is not sufficiently discussed in terms of practical impact. Higher-order NCC algorithms lack detailed gate count analysis (acknowledged by authors but limits completeness). Dense notation in some sections may hinder accessibility to broader quantum computing audience. Limited discussion of experimental feasibility and noise robustness for near-term implementations.

Required actions.

1. Complete the truncated text - ensure the full manuscript including the complete proof of Theorem 1 is provided

2. Add more detailed discussion of the practical tradeoffs involving the t sampling overhead
3. Proofread the manuscript for consistency (e.g., 'Department of Computer Science' appears twice in affiliations)

6 Coverage matrix — what each framework caught

All three pipelines run on the same paper text and same backend. The 5-voice peer review shape is shared across QN and ARS; ARC's review is shorter (2 stages: `peer_review` + `quality_gate`) because ARC's design centre is generation, not review. QN adds quantum-CS-specific layers (audit-and-falsify checklist, 11 quantum-CS fallacies, Stage-6 CQE) that neither ARS nor ARC covers; ARS adds dedicated field-analyst + editorial-synthesizer agents; ARC adds a structured-JSON quality verdict with a numeric threshold gate.

Concern	QN	ARS	
Field / discipline identification before review	implicit (skill prompt)	explicit (Phase 0 agent)	
EIC / editor verdict	✓(in <code>review_panel.md</code>)	✓(<code>eic_review_card.md</code>)	
Methodology / statistics review	✓(R1 voice)	✓(<code>methodology_reviewer</code>)	
Domain / literature review	✓(R2 voice)	✓(<code>domain_reviewer</code>)	
Cross-disciplinary / practical impact review	✓(R3 voice)	✓(<code>perspective_reviewer</code>)	
Devil's Advocate / counter-argument review	✓(Devil's Advocate voice)	✓(<code>devils_advocate_reviewer</code>)	
Methodology-evidence consistency check	implicit (in panel prose)	implicit (in methodology reviewer)	✓(A)
Editorial decision letter / revision roadmap	partial (panel verdict)	✓(Phase 2 synthesizer)	
Numeric quality score with threshold gate	6-dim CQE composite (✓)	–	
Audit-and-falsify checklist (deep-research)	✓(QN-specific)	–	
Quantum-CS-specific fallacy taxonomy	✓(QN-specific; 11 fallacies)	–	
Stage-6 6-dim CQE composite	✓(QN-specific)	–	
Token + USD cost ledger per stage	✓(<code>_backend_used.json</code>)	✓(this driver; same shape)	
Structured stage telemetry (pipeline_summary, etc.)	✓(chain emits ARC-style shapes)	–	