Seven Verifier-Certified Improvements on a Published Sparse-SYK Wormhole-Teleportation Instance

Gyalpo Dongo Sebastian Alberdi Juan Angel Luera

August 25, 2026

Abstract

We report a pre-registered, verifier-certified search that produced seven ten-term sparse Sachdev–Ye–Kitaev (SYK) coupling instances on $N = 8$ Majorana fermions, each certified by a frozen, deterministic, offline verifier to be, *simultaneously*, strictly cheaper than the published sparse-SYK instance of Byun, Kim, and Lee [1] (their Eq. 11) and at least its equal on a pre-registered battery of chaos and wormhole diagnostics, on: (a) transpiled two-qubit gate count on a pinned offline IBM FakeMarrakesh backend snapshot (strictly fewer than the baseline's 56 under the identical pinned compilation pipeline); (b) three quantum-chaos diagnostics — per-parity-sector mean level-spacing ratio, gated by Gaussian-orthogonal-ensemble (GOE) distance to the value 0.5307 within a pre-registered per-sector slack of 0.005 (four of the seven use no slack, sitting at least as close to GOE as the baseline in both parity sectors), spectral-form-factor dip-to-plateau strength, and size-winding coherence; and (c) wormhole-teleportation mutual-information asymmetry under both exact dynamics and single-step Trotterization in the same pinned circuit ordering. The teleportation signal is a hard constraint, never an objective, by design: optimizing that signal directly is what produced the fine-tuned, weakly-chaotic instance in the 2022 wormhole-on-a-processor episode. All comparisons here are internal to a single pinned offline pipeline; we present the seven instances as drop-in candidates for the authors' own hardware protocol, not as a device-level result. Every claim in this paper traces to an append-only, hash-chained ledger entry; the verifier reproduces the baseline's published spectral statistics at import time and refuses to arm otherwise, and it provably rejects a pinned integrable decoy for physics reasons. A single eight-island search run examining 17,410 candidates on one laptop sufficed. We also report two structural findings: coupling-sign patterns act as a (near-)pure Majorana gauge for all seven certified metrics, collapsing the 2^{10} sign patterns of a fixed term-set to at most eight equivalence classes; and the Trotterized teleportation asymmetry — not gate count — is the binding constraint, with ledger-recorded candidates transpiling to as few as 38 two-qubit gates (32% below baseline) that fail only that single gate.

1 Introduction

A small, strongly chaotic quantum system that can be run on today's hardware is the working arena for "quantum gravity in the lab" proposals. The Sachdev–Ye–Kitaev model [2, 3] — N Majorana fermions with random all-to-all quartic couplings — is maximally chaotic, holographically dual to near-AdS$_2$ gravity, and supports a concrete laboratory protocol: traversable-wormhole teleportation, in which a negative-energy coupling $e^{i\mu V}$ between two entangled copies renders an operator inserted on the left side recoverable on the right [4, 5, 6]. Sparse variants of SYK retain the chaotic and holographic phenomenology at a fraction of the term count [7, 8], which is what makes hardware demonstrations conceivable.

The cautionary tale is the 2022 demonstration on a Google processor [9]: the nine-term "learned" Hamiltonian used there was obtained by *optimizing the teleportation signal itself*, and subsequent analysis showed the resulting instance was fine-tuned and not convincingly chaotic, with the reported signal an artifact of that fine-tuning [10]. The lesson we adopt as a design principle: the teleportation signal must be a *constraint that a candidate is required to meet*, never the objective being maximized, and chaos must be gated independently.

Recently, Byun, Kim, and Lee [1] published a specific ten-term sparse-SYK instance on $N = 8$ Majoranas (their Eq. 11) together with a hardware demonstration on an IBM Marrakesh-class device. This gives, for the first time in this niche, a concrete published *incumbent*: a single explicit coupling instance with published spectral statistics ($r_{\text{odd}} = 0.537$, $r_{\text{even}} = 0.479$ per parity sector) and a published circuit cost. The question this campaign posed is narrow and falsifiable:

> *Does there exist a ten-term sparse-SYK instance on eight Majoranas (signs ± 1, fixed magnitude) that beats the published Eq. 11 instance simultaneously on transpiled two-qubit gate count, on all three chaos diagnostics, and on both ideal and Trotterized wormhole-teleportation signal under the same pinned circuit ordering — with the whole claim decided by a deterministic offline verifier in under a second?*

The answer is yes, seven times over, from one pre-registered search run ([SEARCH-1]). Section 2 specifies the frozen contract and verifier; Section 3 the search strategy; Section 4 the certified hits, the near-miss structure, and the cost of the refuted move classes; Section 5 what the run revealed about the landscape. Every number in this paper is either a frozen verifier constant, a value the verifier reproduced at dress rehearsal, or a value recorded verbatim in the campaign's append-only ledger; the provenance tag accompanies each claim.

2 The search contract and the frozen verifier

2.1 Candidate space

A candidate is a pair $(\mathcal{T}, \mathcal{S})$: a set $\mathcal{T}$ of exactly ten distinct 4-subsets of $\{1, \dots, 8\}$ (each written strictly increasing — unsorted quartic Majorana terms are sign-ambiguous) and a sign assignment $\mathcal{S} \in \{\pm 1\}^{10}$. Every term has fixed magnitude $|J| = \sqrt{2}/\sqrt{10}$, so the Hamiltonian is

$$H \;=\; \sum_{a=1}^{10} s_a \, J \, i^2 \, \psi^{j_1(a)} \psi^{j_2(a)} \psi^{j_3(a)} \psi^{j_4(a)}, \qquad s_a \in \{\pm 1\}, \tag{1}$$

with the standard Hermitian quartic convention fixed in the verifier source. There are $\binom{70}{10} \approx 4 \times 10^{11}$ term-sets, each with 2^{10} sign patterns; exhaustion is impossible and was pre-registered as *not* the strategy. The verifier excludes the pinned baseline and its entire 8! Majorana-relabeling orbit (with parity-corrected signs), so a hit is never a disguised copy of the incumbent.

2.2 The seven gates

`verify(candidate)` evaluates, with no short-circuiting past format and orbit exclusion (so near-misses are ledger-reconstructable), all seven gates of Table 1; `valid=True` iff every gate passes. The gates:

1. **Transpiled two-qubit count (strict).** One `PauliEvolutionGate` per term in a pinned deterministic greedy commuting-group ordering, transpiled to the offline `FakeMarrakesh` snapshot at `qiskit==2.5.2` [11], `seed_transpiler=7`, `optimization_level=3`. The candidate's two-qubit operation count must be *strictly below* the baseline's count of 56 under this identical pipeline.

Table 1: The seven gates and the pinned baseline (Byun et al. Eq. 11) values they are measured against. All values are internal to the frozen verifier's pinned pipeline. Provenance: frozen verifier constants and import-time dress rehearsal; the SFF, size-winding, A_{ideal} and A_{Trot} baseline values are additionally recorded verbatim in near-miss margin entries of the search ledger [SEARCH-1:near-miss].

Gate	Baseline value	Candidate passes iff
Transpiled 2q count	56	< 56 (strict)
r_{odd}	0.5372 (GOE dist. 0.0065)	$\lvert r_{\text{odd}} - 0.5307\rvert \le 0.0065 + 0.005$
r_{even}	0.4790 (GOE dist. 0.0517)	$\lvert r_{\text{even}} - 0.5307\rvert \le 0.0517 + 0.005$
SFF dip→plateau strength	215.08	≥ 215.08
Size-winding coherence	0.6998	≥ 0.6998
A_{ideal}	0.011912	≥ 0.011912
A_{Trot}	0.019953	≥ 0.019953
Commuting term pairs	34 of 45	(diagnostic only, not a gate)

2. **Chaos consensus (three gates, never a weighted sum).** (i,ii) Per-parity-sector mean consecutive-level-spacing ratios $r_{\text{odd}}, r_{\text{even}}$ [12, 13], gated on *closeness to the GOE value* 0.5307: per sector, $|r - 0.5307| \le |r_{\text{base}} - 0.5307| + \varepsilon_r$ with pre-registered slack $\varepsilon_r = 0.005$. Raw r inflating past the GOE value is picket-fence level repulsion, not stronger chaos; the campaign's own pre-freeze review caught an early draft gating on raw $r \ge r_{\text{base}}$, which had rejected a candidate at GOE-distance 0.0031 against a baseline at distance 0.0065 — the more chaotic instance failing the chaos gate. The gate was corrected and re-frozen before any search. (iii) Spectral form factor [14] dip-to-plateau strength (plateau over dip on a fixed 241-point logarithmic time grid) $\ge$ baseline, and (iv) size-winding coherence [6] at $\beta = 3$, $t = 1.8$, $\ge$ baseline.

3. **Teleportation constraint (two gates).** Mutual-information asymmetry

$$A = \max_{t_1 \in \{0, 0.1, \ldots, 6\}} \big[I_{\mu=-12}(t_1) - I_{\mu=+12}(t_1)\big], \qquad V = \tfrac{i}{32} \sum_{j=1}^{8} \psi_L^j \psi_R^j, \tag{2}$$

at $\beta = 3$, $t_0 = 1.8$, computed *both* with exact dynamics (A_{ideal}) *and* with the single-step Trotter product [15] in the *same* pinned greedy ordering used for the transpiled circuit (A_{Trot}); each must be $\ge$ baseline. This pairing is the anti-fine-tuning device: an instance whose signal survives only in exact dynamics, or only in some favorable ordering, cannot pass.

2.3 Self-validation: dress rehearsal and integrable decoy

The verifier is *unarmed by construction* until an import-time dress rehearsal passes. It must reproduce the baseline's published per-sector spacing ratios ($r_{\text{odd}} = 0.537$, $r_{\text{even}} = 0.479$, tolerance 2×10^{-3}), the 34/11 commuting/non-commuting pair split, the pinned transpiled count of 56, and positive teleportation asymmetry both ideal and Trotterized — guarding against a silently mistranscribed baseline or environment drift. It must additionally *reject* a pinned fully-commuting integrable control instance (45/45 commuting pairs, massively degenerate spectrum) for physics reasons — mean-r GOE-distance collapse in both parity sectors — guarding against a false-positive-factory

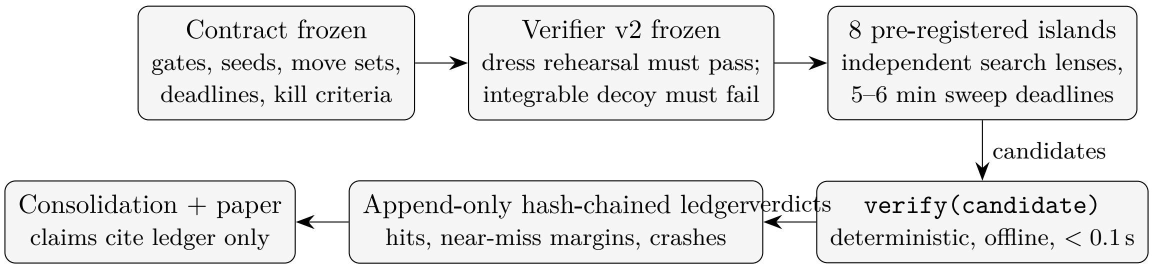

Figure 1: The campaign loop. Everything left of the searcher islands is frozen before search begins; everything the searchers produce is filtered through the frozen verifier before it can enter the record.

verifier. Both checks passed at freeze [SEARCH-1:premortem]; a searcher chasing low gate count is naturally drawn toward highly-commuting term-sets, which is exactly the failure mode this decoy patrols.

2.4 Harness protocol

The campaign ran inside an autonomous research harness whose protocol is designed so that no number in this paper is self-reported:

- **Pre-registration.** The contract — candidate space, all seven gate definitions, seed islands, per-island move classes, sweep deadlines, kill criteria, and a hard cap of 30 search runs — was frozen, and the verifier reviewed and locked, *before* the first candidate was evaluated. A pre-freeze adversarial review round corrected the chaos-gate physics (raw-r to GOE-distance, §2.2) and was itself recorded.
- **Frozen deterministic oracle.** `verify()` is deterministic, fully offline, and evaluates in $\sim$0.07–0.1 s per candidate after import. The claim "hit" is defined as `valid=True` from this frozen code path, nothing else.
- **Append-only, hash-chained ledger.** Every island report, every hit, every near-miss (with its failing gate and numeric margin), and every crash is a ledger entry carrying the hash of its predecessor. Searcher sessions are disposable; only verifier verdicts enter the record.
- **Ship rule.** Pre-registered: any single `valid=True` candidate ships the campaign. No cherry-picking is possible because there is no post-hoc metric aggregation — the gates are a hard consensus fixed in advance.

Figure 1 shows the loop. This campaign had *zero* iterative optimize-and-promote cycles (harness stats: cycles = 0, experiments = 0, promotions = 0); all results below come from the single search run's ledger.

3 Search strategy

The predecessor campaign (`syk-wormhole-beat`, verifier v1: raw-r chaos gate, no Trotterized-asymmetry gate) had recorded broad enumeration of structured families as dead, five v1-valid candidates, and 671 unique near-misses. Its "signal champion" — which transpiles to 49 two-qubit gates under the v2 pipeline and passes six of the seven v2 gates, failing only the new A_{Trot} gate by 0.016650 vs. 0.019953 [SEARCH-1:near-miss] — was the principal seed. The pre-registered strategy

Table 2: The seven certified hits (full specifications in Appendix A) and the ledger-recorded discovery move for each. "Champion" is the predecessor campaign's v1 signal champion (Appendix A); hits 6 and 7 share one term-set.

Hit	Island (lens)	Discovery move (as recorded)
1	2 (structure-first)	Single-term swaps off the champion after gauge-coset reduction of the sign space ($2^{10} \to 8$ classes).
2	4 (random-wide)	Uniform random 10-subset with uniform random signs, seed 20260825; hit at ∼sample 1,000 of 3,342.
3	5 (near-miss mining)	Second-closest predecessor near-miss ($A_{\text{Trot}} = 0.0133$) with term $[2, 4, 7, 8]$ swapped for $[2, 5, 6, 7]$.
4	6 (adversarial)	Two-term swap off the champion (drop $[1, 2, 5, 6]$, $[1, 4, 5, 6]$; add $[1, 3, 6, 8]$, $[2, 4, 6, 8]$) with sign re-draw.
5	7 (composite)	Three-term swap grafting baseline-side terms ($[1, 3, 7, 8]$, $[1, 5, 6, 7]$, $[2, 3, 5, 7]$) onto the champion's low-gate-count structure.
6	7 (composite)	Three-term swap (kept 7 champion terms; imported $[1, 5, 6, 7]$, $[2, 3, 5, 7]$, $[2, 4, 6, 7]$) plus a single sign flip.
7	7 (composite)	Same term-set as hit 6, different single sign flip.

was local perturbation around recorded survivors: exhaustive sign sweeps (2^{10} per term-set), single-term swaps (one 4-subset replaced by one of the ∼60 unused ones), and two-term swaps only after one-swap exhaustion.

Eight islands ran in parallel with pre-registered, deliberately diverse lenses (Table 3): small-moves-first, structure-first (gauge analysis), boundary probing, uniform random sampling as a control, predecessor near-miss mining, an adversarial lens instructed to search where the architect's prescriptions were vaguest, a composite lens blending partial solutions, and a literature lens (Steiner quadruple systems, cyclic difference designs, hypercube dimer-plaquette cores). A first launch of the run crashed on all eight islands before evaluating any candidate (external code-execution provider rate limit); it is recorded in the ledger and was not counted against the run budget [SEARCH-1:crash].

4 Results

4.1 Seven certified hits

Run 1 explored 17,410 candidates across the eight islands within their 5–6 minute sweep deadlines; seven candidates returned `valid=True` from the frozen verifier and 23 unique near-misses were recorded with their failing margins [SEARCH-1:run]. Table 2 summarizes the hits and the move that found each; Appendix A lists all seven certified instances verbatim. The seven hits span six distinct term-sets (hits 6 and 7 share a term-set and differ in signs). Each hit's certificate is the frozen verifier's verdict that *all seven* gates of Table 1 pass simultaneously: strictly fewer than 56 transpiled two-qubit gates, both parity sectors within the GOE-distance envelope, SFF strength ≥ 215.08, size winding ≥ 0.6998, $A_{\text{ideal}} \geq 0.011912$, and $A_{\text{Trot}} \geq 0.019953$ [SEARCH-1:hit-0..6].

4.2 The binding constraint is the Trotterized signal, not the gate count

Of the 23 recorded near-misses, 10 fail *only* the A_{Trot} gate (Figure 3). The ledger records their exact margins: candidates transpiling to 38, 40, 40, 46, 47, 49 and 51 two-qubit gates — as much as 32% below the baseline's 56 — pass all six other gates and miss only on A_{Trot}, with values between

Table 3: Search run 1 by island [SEARCH-1:island-1..8]. Islands reported 26 near-miss candidates; 23 unique entries remain in the ledger after deduplication. Total candidates: 17,410.

Island	Lens	Candidates	Hits	Near-misses
1	small-first	3,249	0	3
2	structure-first	477	1	2
3	boundary	389	0	2
4	random-wide	3,342	1	5
5	near-miss mining	2,158	1	3
6	adversarial	4,200	1	5
7	composite	3,454	3	3
8	literature designs	141	0	3
Total		17,410	7	26 (23 unique)

0.00628 and 0.01665 against the required 0.019953 [SEARCH-1:near-miss]. Figure 2 plots this frontier. Cheap circuits with baseline-grade chaos and ideal-dynamics signal are abundant; cheap circuits whose signal *survives single-step Trotterization in the pinned ordering* are the scarce commodity. This is the quantitative shape of the 2022 lesson [9, 10]: the Trotterized-signal constraint is doing real exclusion work precisely where an objective-driven search would have declared victory.

4.3 Sign patterns are (near-)pure gauge

The structure-first island proved, and three exhaustive sweeps confirmed empirically, that the Majorana reflection $\psi^j \to -\psi^j$ — a unitary conjugation that flips the sign of every term containing index j — leaves all seven certified metrics exactly invariant. Consequently the $2^{10} = 1024$ sign patterns of a fixed term-set collapse to at most $2^{10}/2^7 = 8$ gauge cosets. The ledger records the sharpest version of this: across the champion term-set's full 1024-pattern sweep, A_{Trot} is pinned at 0.016650 for *every* pattern, and across the $[2, 5, 6, 7]$-swap term-set, r_{odd} is pinned at 0.54268 for every pattern [SEARCH-1:run]. Two refutations follow at recorded cost: no sign pattern on the champion term-set can ever pass v2 (1,024 candidates), and none on the $[2, 5, 6, 7]$-swap term-set (1,024 candidates). Sign-only sweeps of a fixed term-set are a dead move class; the search-relevant object is the term-set (with signs a cheap eight-coset formality), which is why every hit in Table 2 was found by a term-set move.

4.4 Refuted move classes and their cost

Beyond the two sign-sweep refutations above, the full one-swap neighborhood of the champion — all 10 positions × 60 replacement subsets × both signs for the new term, 1,200 candidates — contains zero hits; only 22 of them fail exactly one gate [SEARCH-1:run]. The closest miss in the entire run is instructive: the champion with $[1, 4, 6, 8] \to [2, 5, 6, 7]$ passes six gates including A_{Trot} and misses the odd-sector mean-r gate by 4.2×10^{-4} in GOE-distance (0.01198 vs. allowed 0.01156) [SEARCH-1:near-miss]. The literature-design island (Steiner quadruple systems, cyclic difference sets, hypercube dimer cores; 141 candidates) produced no hits — highly structured, highly commuting term-sets are exactly what the chaos gates punish, consistent with the integrable-decoy physics of §2.3.

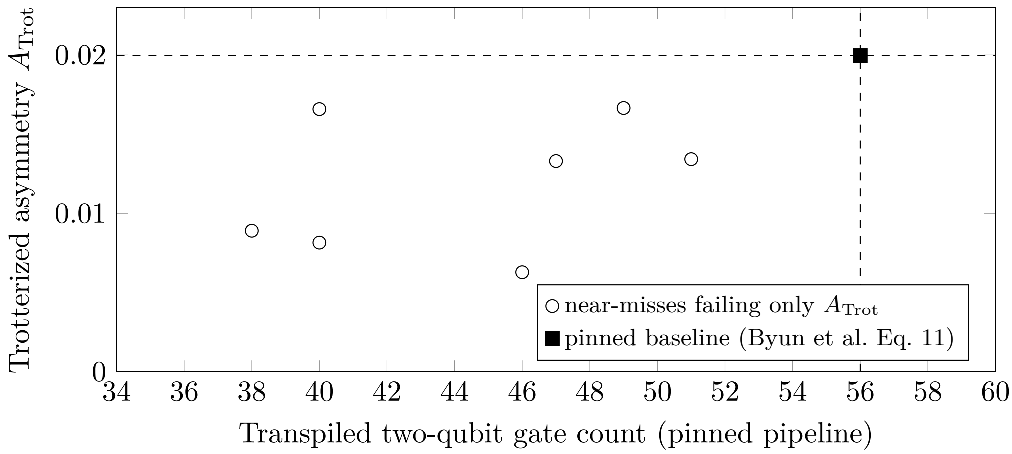

Figure 2: The feasibility frontier from ledger-recorded one-gate near-misses [SEARCH-1:near-miss]. A hit must lie strictly left of the vertical dashed line (fewer than 56 two-qubit gates) *and* on or above the horizontal dashed line ($A_{\mathrm{Trot}} \geq 0.019953$), while also passing the four chaos/ideal-signal gates. The plotted circles pass all gates except A_{Trot}; the seven certified hits occupy the upper-left region by certification (their full numeric transcripts are recomputed deterministically by the reproduction package, §8).

4.5 Cost

The entire evidence base of this paper is one search run: 17,410 verifier calls at $\sim$0.07–0.1 s each, eight islands within 5–6 minute sweep deadlines, on a single laptop (Apple-silicon MacBook Pro), plus one aborted launch that consumed no verifier calls. Verifier freeze, including the adversarial review round that corrected the chaos-gate physics, took under two hours of wall clock on the same day [SEARCH-1:journal].

5 Discussion

Substantial headroom remains under the incumbent's own success criteria. Seven certified instances improve on the published Eq. 11 instance simultaneously on cost, chaos, and both signal gates, and were found in minutes by pre-registered local search. The near-miss frontier (Figure 2) suggests substantial remaining headroom on gate count at fixed signal.

Uniform random sampling is not dead under the corrected chaos gate. The predecessor campaign recorded broad enumeration as dead under verifier v1's raw-r gate. Under v2's GOE-distance gate — the physically correct formulation — a pure uniform-random island found a certified hit within $\sim$1,000 samples [SEARCH-1:hit-1]. The feasible set is not measure-tiny; the earlier gate was partly rejecting *more*-chaotic candidates for exceeding the baseline's distance from GOE on the wrong side. Verifier physics, not search cleverness, was a principal bottleneck — a concrete argument for adversarial review of the judge before any search is trusted.

What a hit means physically. Each certified instance is a ten-term sparse-SYK Hamiltonian whose spectrum is baseline-grade chaotic in both parity sectors (by GOE-distance), whose scrambling diagnostics (SFF ramp structure, size winding) are at least baseline, and whose traversable-wormhole mutual-information asymmetry survives single-step Trotterization in the exact ordering that its cheaper-than-baseline circuit compiles — i.e., the signal is a property of the *implementable*

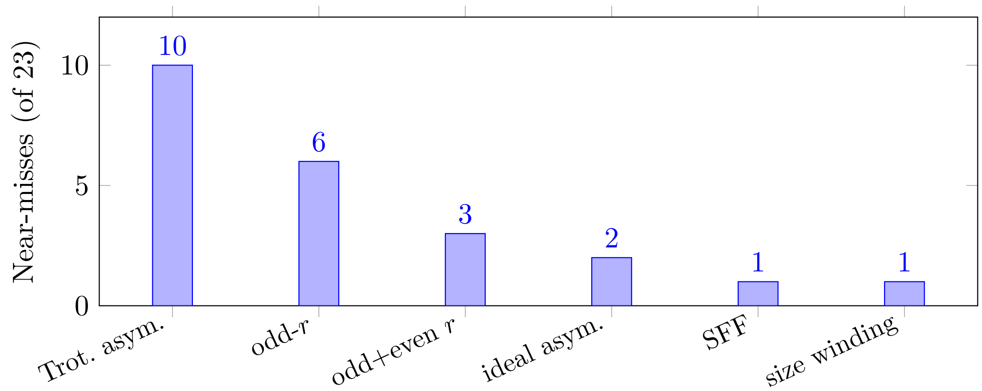

Figure 3: Failing gate across the 23 unique recorded near-misses [SEARCH-1:near-miss]. The Trotterized teleportation asymmetry is the dominant point of failure; the two mean-r categories are near-misses failing the odd sector only vs. both parity sectors.

circuit, not only of idealized dynamics. These are, by construction, better hardware targets than the published instance under the published instance's own figure-of-merit family.

6 Limitations

- **Pipeline-internal comparison.** All gate counts are internal to one pinned pipeline (greedy commuting-group ordering, `qiskit==2.5.2`, `optimization_level=3`, `seed_transpiler=7`, offline FakeMarrakesh snapshot), under which the baseline compiles to 56 two-qubit gates. The baseline paper's own compiler reports a different hardware count for its full protocol; we make no cross-compiler claim. The campaign contract's prose quoted the predecessor pipeline's baseline count of 60; the frozen verifier's own dress-rehearsal-pinned count of 56 is the stricter bar actually enforced, and all seven hits clear it.

- **No hardware execution.** All dynamics are noiseless statevector computations at $N = 8$ per side. Device noise, which the gate-count reduction directly mitigates, is otherwise unmodeled.

- **Single-step Trotterization only.** A_{Trot} certifies one Trotter step in one pinned ordering — the anti-fine-tuning device of this contract, not a convergence study.

- **Relative, not absolute, chaos.** The chaos gates are relative to the baseline with pre-registered slack $\varepsilon_r = 0.005$. The baseline's even sector ($r_{\mathrm{even}} = 0.479$, GOE-distance 0.0517) is itself not GOE-converged at this system size, so "baseline-grade" is the honest reading throughout.

- **Provenance of the per-hit numbers.** The full seven-gate transcript of every certified hit (Table 4) is recomputed bit-for-bit by the frozen verifier and recorded verbatim in the ledger consolidation row [SEARCH-1:consolidation-transcripts]; the reproduction package (§8, `reproduce.py`) regenerates it in under a second per candidate. The chaos slack is used by three hits in one sector each (Table 4, †); the other four sit at least as close to GOE as the baseline in both sectors.

- **Seven candidates, six term-sets.** Hits 6 and 7 share a term-set and differ by sign flips;

given the observed gauge structure they may be gauge-equivalent (the ledger does not adjudicate this). The conservative count is six distinct instances up to signs.

- **Harness-build provenance.** Eight ledger rows of search run 1 were recorded under a harness build (`harness_sha 95d62cf`) later than the reference build (`8cb3da7`); the difference (commit `b492153`) is re-verification plumbing only. The frozen verifier's `contract_sha` is unchanged across both builds, so no measured value depends on the drift; we record it for completeness.

- **Contract prose vs. frozen verifier.** The frozen `CONTRACT.md` prose predates two in-place corrections and still describes the chaos gate via raw sector-averaged $\langle r \rangle$ (superseded by the enforced GOE-distance gate) and quotes the predecessor pipeline's baseline gate count of 60 (superseded by the dress-rehearsal-pinned 56). In both cases the binding criterion is the one compiled into the frozen `verifier.py` and exercised at the import-time dress rehearsal — pinned by the verifier's `contract_sha` — not the contract prose, which is left unedited to preserve the freeze.

- **Baseline citation.** The baseline is the sparse-SYK instance of Byun, Kim, and Lee [1] (their Eq. 11), pinned by arXiv identifier and equation number in the frozen contract before any search. The full bibliographic record is given in the references.

7 AI-contribution statement

The authors are fully responsible for the content of this paper. This campaign was executed inside an autonomous multi-agent research harness: AI systems (Anthropic Claude for planning, contract drafting, adversarial verifier review, and prose; OpenAI Codex for search-island code execution) generated and swept the candidate Hamiltonians and drafted this manuscript. No AI system is an author. The scientific safeguards are mechanical rather than trust-based: the success criterion was frozen as a deterministic offline verifier before any search; the verifier self-validates against the published baseline values and a pinned integrable decoy at import time; every claimed number traces to an append-only, hash-chained ledger of verifier verdicts; and no self-reported model output enters the record. The human authors reviewed and approved the contract and verifier before freeze, reviewed the ledger, and take responsibility for verification of the results and for every citation in the bibliography.

8 Code availability

The claims in this paper are checkable without any part of our search harness. Reproduction requires only a small self-contained package: the frozen contract, the frozen verifier (`verifier.py`, including its import-time dress rehearsal against the Byun–Kim–Lee published spectral statistics and its integrable-decoy check), the seven certified instances and all recorded near-misses with their verdicts, the append-only ledger of search run 1, and the environment pins (`qiskit==2.5.2`, the offline FakeMarrakesh snapshot version) needed to recompute every gate value deterministically in under a second per candidate. This package was built and run locally; it is not currently published, and is available from the authors on request. The multi-agent search harness that discovered the instances is proprietary and is deliberately *not* part of the package — and is not needed to check the result: verifying any claim here means re-running the frozen verifier on the instances (the package's `reproduce.py` does exactly this, regenerating Table 4 and re-confirming the baseline's exclusion),

Instance	2q	$\langle r\rangle_{\text{odd}}$	$\langle r\rangle_{\text{even}}$	SFF	S_{wind}	$\mathcal{A}_{\text{ideal}}$	$\mathcal{A}_{\text{trot}}$
Byun et al. Eq. 11	56	0.5372	0.4788	215.1	0.700	0.01191	0.01995
Hit 1	41	0.5307	0.5129	262.6	0.933	0.01259	0.02576
Hit 2	53	0.5270	0.4932	281.7	0.734	0.01298	0.03297
Hit 3	45	0.5277	0.5590	227.0	0.720	0.01331	0.02851
Hit 4	51	0.5296	0.5072	470.1	0.937	0.01546	0.02798
Hit 5	40	0.5353	$0.5828^{\dagger}$	928.4	0.977	0.04080	0.03426
Hit 6	53	$0.5209^{\dagger}$	0.5774	524.8	0.912	0.01230	0.04669
Hit 7	53	$0.5209^{\dagger}$	0.5774	524.8	0.983	0.01517	0.03039

Table 4: The full seven-gate transcript of every certified instance versus the pinned baseline, recomputed bit-for-bit by the frozen verifier and recorded in the ledger ([SEARCH-1:consolidation-transcripts]). Columns: transpiled two-qubit gate count on the offline FakeMarrakesh snapshot; per-parity-sector mean level-spacing ratio $\langle r\rangle$ (chaos gate is per-sector GOE-distance $|\langle r\rangle - 0.5307| \leq |\langle r\rangle_{\text{base}} - 0.5307| + 0.005$); spectral-form-factor dip-to-plateau strength; size-winding coherence; and ideal / single-step-Trotterized teleportation asymmetry. Every hit has strictly fewer two-qubit gates and strictly larger SFF, S_{wind}, $\mathcal{A}_{\text{ideal}}$ and $\mathcal{A}_{\text{trot}}$ than the baseline. † marks the one sector in which Hits 5–7 sit farther from GOE than the baseline, within the pre-registered 0.005 slack; Hits 1–4 are at least as close to GOE as the baseline in *both* sectors (no slack used).

not trusting or inspecting the process that found them — which is precisely the guarantee the frozen-verifier design is built to provide.

A The seven certified instances

Each instance is $\{\mathcal{T}, \mathcal{S}\}$ with $|J| = \sqrt{2}/\sqrt{10}$ per term; terms are 4-subsets of $\{1, \ldots, 8\}$, written strictly increasing, with their signs listed in the same order. Verbatim from the ledger [SEARCH-1:hit-0..6].

Hit 1 terms $[1,2,3,7]$, $[1,2,5,6]$, $[1,3,4,5]$, $[1,3,4,7]$, $[1,4,5,6]$, $[1,4,6,8]$, $[2,4,5,8]$, $[2,4,7,8]$, $[3,4,5,6]$, $[4,6,7,8]$;
signs $-,-,-,-,+,-,+,+,+,+$.

Hit 2 terms $[1,2,3,4]$, $[1,2,3,7]$, $[1,2,3,8]$, $[1,3,4,7]$, $[1,4,5,7]$, $[1,4,5,8]$, $[2,4,6,8]$, $[3,4,7,8]$, $[3,5,6,8]$, $[3,5,7,8]$;
signs $-,+,+,+,+,-,-,-,+,+$.

Hit 3 terms $[1,3,4,8]$, $[1,4,6,8]$, $[1,5,6,7]$, $[2,3,7,8]$, $[2,5,6,7]$, $[3,4,5,7]$, $[3,5,6,7]$, $[3,5,6,8]$, $[3,5,7,8]$, $[4,5,7,8]$;
signs $-,-,-,+,+,+,+,+,-,-$.

Hit 4 terms $[1,2,3,7]$, $[1,3,4,5]$, $[1,3,4,7]$, $[1,3,6,8]$, $[1,4,6,8]$, $[1,5,7,8]$, $[2,4,6,8]$, $[2,4,7,8]$, $[3,4,5,6]$, $[4,6,7,8]$;
signs $-,-,-,+,+,-,+,+,+,+$.

Hit 5 terms $[1,2,5,6]$, $[1,3,4,5]$, $[1,3,4,7]$, $[1,3,7,8]$, $[1,4,5,6]$, $[1,4,6,8]$, $[1,5,6,7]$, $[1,5,7,8]$, $[2,3,5,7]$, $[2,4,7,8]$;
signs $-,-,-,+,+,-,+,-,+,+$.

Hit 6 terms $[1, 2, 5, 6]$, $[1, 3, 4, 7]$, $[1, 4, 5, 6]$, $[1, 4, 6, 8]$, $[1, 5, 6, 7]$, $[1, 5, 7, 8]$, $[2, 3, 5, 7]$, $[2, 4, 6, 7]$, $[2, 4, 7, 8]$, $[4, 6, 7, 8]$;
signs $-, -, +, +, +, +, +, -, +, +$.

Hit 7 same terms as Hit 6;
signs $-, -, +, +, +, -, +, -, -, +$.

For reference, the pinned baseline (Byun et al. Eq. 11, excluded together with its full 8! relabeling orbit): terms $[1, 2, 4, 7]$, $[1, 2, 6, 8]$, $[1, 3, 4, 5]$, $[1, 3, 7, 8]$, $[1, 4, 5, 6]$, $[1, 5, 6, 7]$, $[2, 3, 4, 8]$, $[2, 3, 5, 7]$, $[2, 4, 6, 7]$, $[3, 5, 6, 8]$; signs $-, +, -, +, -, +, -, +, -, +$. The predecessor v1 signal champion (a v2 near-miss, $A_{\text{Trot}} = 0.016650$, transpiled count 49): terms $[1, 2, 3, 7]$, $[1, 2, 5, 6]$, $[1, 3, 4, 5]$, $[1, 3, 4, 7]$, $[1, 4, 5, 6]$, $[1, 5, 7, 8]$, $[1, 4, 6, 8]$, $[2, 4, 7, 8]$, $[3, 4, 5, 6]$, $[4, 6, 7, 8]$; signs $-, -, -, -, +, -, +, +, +, +$.

References

[1] Moongul Byun, Keun-Young Kim, and Hyeonsoo Lee. Quantum simulation of traversable-wormhole-inspired quantum teleportation in a chaotic binary sparse SYK model, 2026. arXiv:2604.10090. Eq. 11 is the pinned baseline instance of the present study.

[2] Subir Sachdev and Jinwu Ye. Gapless spin-fluid ground state in a random quantum Heisenberg magnet. *Phys. Rev. Lett.*, 70:3339–3342, 1993.

[3] Juan Maldacena and Douglas Stanford. Remarks on the Sachdev–Ye–Kitaev model. *Phys. Rev. D*, 94:106002, 2016. arXiv:1604.07818.

[4] Ping Gao, Daniel Louis Jafferis, and Aron C. Wall. Traversable wormholes via a double trace deformation. *J. High Energy Phys.*, 12:151, 2017. arXiv:1608.05687.

[5] Juan Maldacena and Xiao-Liang Qi. Eternal traversable wormhole, 2018. arXiv:1804.00491.

[6] Adam R. Brown, Hrant Gharibyan, Stefan Leichenauer, et al. Quantum gravity in the lab: teleportation by size and traversable wormholes, 2019. arXiv:1911.06314.

[7] Shenglong Xu, Leonard Susskind, Yuan Su, and Brian Swingle. A sparse model of quantum holography, 2020. arXiv:2008.02303.

[8] Antonio M. García-García, Yiyang Jia, Dario Rosa, and Jacobus J. M. Verbaarschot. Sparse Sachdev–Ye–Kitaev model, quantum chaos, and gravity duals, 2020. arXiv:2007.13837.

[9] Daniel Jafferis, Alexander Zlokapa, Joseph D. Lykken, David K. Kolchmeyer, Samantha I. Davis, Nikolai Lauk, Hartmut Neven, and Maria Spiropulu. Traversable wormhole dynamics on a quantum processor. *Nature*, 612:51–55, 2022.

[10] Bryce Kobrin, Thomas Schuster, and Norman Y. Yao. Comment on “Traversable wormhole dynamics on a quantum processor”, 2023. arXiv:2302.07897.

[11] Qiskit contributors. Qiskit: an open-source framework for quantum computing, 2023. DOI 10.5281/zenodo.2573505; version 2.5.2 pinned in this work.

[12] Vadim Oganesyan and David A. Huse. Localization of interacting fermions at high temperature. *Phys. Rev. B*, 75:155111, 2007.

[13] Y. Y. Atas, E. Bogomolny, O. Giraud, and G. Roux. Distribution of the ratio of consecutive level spacings in random matrix ensembles. *Phys. Rev. Lett.*, 110:084101, 2013.

[14] Jordan S. Cotler, Guy Gur-Ari, Masanori Hanada, Joseph Polchinski, Phil Saad, Stephen H. Shenker, Douglas Stanford, Alexandre Streicher, and Masaki Tezuka. Black holes and random matrices. *J. High Energy Phys.*, 05:118, 2017. arXiv:1611.04650.

[15] Seth Lloyd. Universal quantum simulators. *Science*, 273:1073–1078, 1996.