QuantumNovelty Review Showcase

Paper audit 20 Second Parity Lifetime in an InAs-Pb Tetron Device (arXiv:2606.03884)

Generated by the QuantumNovelty paper-audit pipeline

June 2026

Project	QuantumNovelty (QN) — audit-and-falsify framework for quantum-computing research
Repository	`https://github.com/boltzmannentropy/QuantumNovelty`
Author	Shlomo Kashani (QNeura.ai)
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 23:18 by `build_reviews.py`

Disclaimer. These reviews are generated end-to-end by AI (the QuantumNovelty agent pipeline running on the Claude Code CLI). They are provided strictly for academic and demonstration purposes — to show what the framework produces on real papers. We make NO claims about the correctness, quality, novelty, or publication-worthiness of the papers under review; four passed real peer review at their respective journals and one is a public arXiv preprint, and nothing here should be read as criticism of the authors or as a substitute for human peer review. Paper copyrights remain with their authors and publishers.

1 Papers under review

Tag	Paper	Venue	arXiv
`majorana2`	20 Second Parity Lifetime in an InAs-Pb Tetron Device Microsoft Quantum	arXiv:2606.03884 (preprint, June 2026)	2606.03884

2 Token + cost ledger

Every LLM call records the model snapshot ID, exact input/output token counts, and USD cost from the Claude CLI's JSON envelope.

Paper	Stage	Input tk	Output tk	Cost ($)	Elapsed (s)
majorana2	Deep-research review	1	1,451	$0.2954	39.8
majorana2	Reviewer panel 5 voices	1	5,072	$0.3114	121.7
majorana2	Logical-fallacy report	1	3,112	$0.2598	62.7
majorana2	Argument-structure audit	1	3,083	$0.2662	73.5
majorana2	Numeric-claim registry (deterministic)	-	-	(est.)	0.0
majorana2	Disclosure audit	1	1,781	$0.2270	33.7
majorana2	Anchored revision plan	2	2,676	$0.3423	56.1
majorana2	Stage-6 CQE narrative	3,005	1,089	$0.0763	28.3
majorana2 total		**3,012**	**18,264**	**$1.7784**	**415.8**
Grand total		**3,012**	**18,264**	**$1.7784**	**415.8**

3 Verdict summary

Paper	Panel score	EIC verdict	CQE composite
`majorana2`	6.33/10	major-revisions	23/100

Panel score is the mean of the five voices' *Verdict: N/10* scores; the EIC verdict comes from the vote table; the CQE composite is the geometric mean of the six process-quality dimensions.

4 majorana2 — *20 Second Parity Lifetime in an InAs-Pb Tetron Device*

Microsoft Quantum arXiv:2606.03884 arXiv:2606.03884 (preprint, June 2026)

4.1 majorana2 — Deep-research review

Model: `claude-opus-4-5-20251101` · 1,451 output tokens · 39.8 s

4.1.1 1. One-paragraph summary of what the paper claims

This paper from Microsoft Quantum reports a major advance in topological quantum computing by demonstrating a characteristic parity lifetime of approximately 20 seconds in an InAs-Pb tetron device, representing an improvement of more than three orders of magnitude over previous Al-based devices (which achieved 1–12 ms). The authors attribute this improvement to replacing aluminum with the higher-gap superconductor lead (Pb) in their superconductor-semiconductor hybrid platform, which yields a larger topological gap (T 70 ţeV vs. ~30 ţeV in Al devices) and a proximity-induced gap of ~570 ţeV. They introduce an rf-based wire spectroscopy technique for scalable device bring-up that can resolve Majorana hybridization energies (EM) with ţeV precision, and demonstrate h/2e-periodic bimodal quantum capacitance shifts consistent with parity-dependent interference. The paper argues this validates a central premise of topological quantum computing: that increasing the excitation gap dramatically reduces error mechanisms and improves qubit performance.

4.1.2 2. Audit-and-falsify checklist

Item	Status	Evidence
Augmented baseline catalog	**PASS**	The paper explicitly compares against their own prior Al-InAs devices (Refs. 59–61) with quantitative improvements cited: parity lifetime improvement from 1–12 ms to ~20 s; topological gap from ~30 ţeV to ~70 ţeV; topological phase region more than doubled; this represents comparison against current state-of-the-art from the same research group.
Strict-domination comparator	**PARTIAL**	The paper reports “more than three orders of magnitude” improvement in parity lifetime and provides specific values (22 ś 1 s fit), but comparisons like “top quintile gap” thresholds and localization length claims lack explicit tolerance specifications (_abs, _rel).
Recompute-from-raw	**PARTIAL**	The paper presents raw data in figures (e.g., time traces in Fig. 6, phase diagrams in Fig. 3) and reports derived values (Z = 22 ś 1 s from N = 324 dwell intervals), but there is no explicit statement that readers can access raw data or re-derive the numerical claims independently.
Wilson 95% CIs	**PARTIAL**	The parity lifetime Z = 22 ś 1 s includes uncertainty from exponential fitting, but the paper does not explicitly provide binomial/Wilson confidence intervals for small-sample statistics; the N = 324 events is reasonably large but the methodology for uncertainty propagation is not fully detailed.
Cross-LLM falsifiability	**NOT-APPLICABLE**	No LLM-in-the-loop methods were used in this experimental physics paper.

Item	Status	Evidence
Honest negatives	**PARTIAL**	The paper acknowledges that “external quasiparticle poisoning events” cause sign switches in their data (Sec. 3–4), and discusses that non-equilibrium quasiparticles can degrade protection, but there is no dedicated “Failure Modes” section systematically cataloging regimes where the device underperforms or the method fails.
Simulator precision floor	**PARTIAL**	The paper mentions simulations for induced gap values (“consistent with simulations”) and uses numerical modeling, but does not explicitly state whether float64 reference paths were used or address numerical precision floors in their calculations.
Auditable claims	**FAIL**	There is no mention of a re-runnable script (e.g., `audit_claims.py`), on-disk JSON, or any reproducibility infrastructure that would allow independent derivation of numerical claims from raw data.

4.1.3 3. Overall assessment

This paper represents solid experimental physics work with clear advances over prior results from the same group. The claims are well-grounded in extensive measurements, the methodology (rf-based spectroscopy, parity injection) is clearly described, and the results are internally consistent. However, from a strict research-rigour audit perspective, the paper falls short in several areas: (1) no publicly accessible raw data or reproducibility infrastructure; (2) uncertainty quantification is present but not comprehensively specified (e.g., explicit confidence interval methodology for the 324-event parity lifetime analysis); (3) baseline comparisons, while appropriate, rely heavily on self-citation without independent verification of the comparison metrics; (4) absence of a systematic failure-modes discussion. The paper would benefit from explicit tolerance specifications in performance claims and a data-availability statement with recomputation scripts.

Research rigour score: 6.5/10 — The experimental methodology and results are credible and represent genuine advances, but the paper lacks the reproducibility infrastructure and explicit uncertainty quantification that strict audit standards require.

4.1.4 4. Three highest-leverage improvements

1. **Add data availability and reproducibility infrastructure**: Include a supplementary package with raw measurement data (time traces, conductance sweeps, TGP outputs) in a standard format (HDF5/JSON) and a script that derives every numerical claim (Z, T, EM bounds, periodicity estimates) from this data. This would transform the “Auditable claims” item from FAIL to PASS.

2. **Provide explicit uncertainty methodology and confidence intervals**: For the key Z = 22 ś 1 s result, specify the statistical framework (maximum likelihood estimation on exponential distribution? Bayesian posterior?), report the 95% CI explicitly, and address potential systematic uncertainties (threshold selection sensitivity, drift during long traces). For the EM < 1 ţeV claim, quantify the confidence level and detector resolution floor more precisely.

3. **Include a dedicated “Limitations and Failure Modes” section**: Systematically document parameter regimes where the device does not achieve the reported performance (e.g., regions outside the identified cluster in Fig. 5, conditions under which poisoning rates increase, sensitivity to magnetic field misalignment), and discuss any instances where the rf spectroscopy method failed to identify low-energy states. This honest reporting would significantly strengthen the paper’s credibility for rigorous reviewers.

4.2 majorana2 — Reviewer panel 5 voices

Model: `claude-opus-4-5-20251101` · 5,072 output tokens · 121.7 s

4.2.1 Voice 1 — Reviewer 1 (Physics correctness)

This manuscript reports on an InAs–Pb hybrid device demonstrating parity lifetimes of approximately 20 seconds, which would represent a substantial improvement over the approximately 1–12 millisecond lifetimes observed in prior aluminum-based tetron devices. The physical premise is straightforward and well-motivated: replacing aluminum (300 ţeV) with lead (1.3 meV) should increase both the parent superconducting gap and, consequently, the topological gap, thereby suppressing both thermal excitations and quasiparticle poisoning. The authors claim to have achieved top-quintile topological gaps of T 70 ţeV, compared to T 30 ţeV in earlier Al-based devices. The basic physics underlying these claims is sound: the exponential dependence of Majorana hybridization on L/T, where T scales inversely with the gap in clean systems, combined with stronger electron-phonon coupling in Pb facilitating faster quasiparticle recombination, provides a coherent theoretical framework for the observed improvements.

The Hamiltonian introduced in Equation 1 describes a minimal model coupling a single low-energy wire state to a readout quantum dot. While this model is adequate for qualitative understanding, I have concerns about its application to the quantitative extraction of EM. The authors state that EM is extracted with “ţeV precision,” yet the model assumes a single isolated wire state. In the topological regime, this assumption is appropriate since 1 and 2 represent well-separated Majorana zero modes with exponentially small overlap. However, near phase boundaries or in the presence of disorder, multiple low-energy states (quasi-Majoranas) can contribute. The authors acknowledge this possibility by referencing prior work on quasi-MZMs (Ref. 61, 84), but do not explicitly demonstrate that their measurements exclude such scenarios. The claim that EM < 1 ţeV throughout the identified topological region would be substantially strengthened by showing the parity-independence of peak width and amplitude predicted by Equation 2 in the deep topological limit where tR 0.

The induced gap measurements deserve careful scrutiny. The authors report ind 570 ţeV in the lowest subband regime at zero field for the nanowire geometry, which they claim is consistent with simulations accounting for transverse confinement. However, the proximity-induced gap measured in the 2D geometry (Fig. 2b) is 400 ţeV. While transverse confinement can indeed enhance the induced gap, the 42% increase from 400 to 570 ţeV seems substantial and warrants explicit presentation of the simulation methodology. Additionally, the relationship between the parent Pb gap (Pb 1.3 meV) and the induced gap is governed by interface transparency and the relative density of states in superconductor and semiconductor. The 10 nm Pb film thickness is specified, but the interface characterization that would support claims of “hard” induced gap (absence of subgap states) relies on a citation to prior Coulomb island work (Ref. 53) rather than direct measurement in the present device geometry.

The spin-orbit coupling value 12–16 meV nm is extracted from Shubnikov-de Haas oscillations in Pb-proximitized nanowires according to Fig. 2c. The extraction procedure references Ref. 78, but the actual data shown appears to come from van der Pauw devices with fixed density. The authors acknowledge in their footnote (Ref. 77) that this value combines 2D measurements under Pb with weak anti-localization measurements in Hall bars without Pb. This indirect approach introduces systematic uncertainty that is not propagated into the claimed precision. Given that the topological phase boundary depends sensitively on the ratio of spin-orbit energy to Zeeman energy, more direct characterization of in the actual device geometry would strengthen the conclusions about phase diagram extent.

Questions for Authors:

1. Can you provide explicit numerical simulations demonstrating the 42% enhancement of induced gap from transverse confinement, including the specific model assumptions about interface transparency?
2. In the topological region identified in Fig. 5, do the quantum capacitance peaks exhibit parity-independent width and amplitude as predicted by Equation 2 when tR 0, and if so, can you quantify the upper bound on tR/tR?
3. What is the estimated systematic uncertainty in the spin-orbit coupling arising from the indirect measurement approach combining van der Pauw and Hall bar data?

Verdict: 7/10 — minor-revisions

4.2.2 Voice 2 — Reviewer 2 (Algorithmic novelty)

The central novelty claim of this work is the demonstration that increasing the superconducting gap translates directly into improved topological qubit performance, specifically through the replacement of aluminum with lead in superconductor-semiconductor hybrid devices. While this is indeed a longstanding prediction of topological quantum computing theory, the question of algorithmic or methodological novelty requires careful examination. The rf-based wire spectroscopy technique presented in Section 3 represents a genuine methodological advance over prior DC transport-based tuning protocols, enabling parallel characterization compatible with scalable architectures. This addresses a practical bottleneck identified in earlier work and represents the most novel technical contribution of the manuscript.

Comparing against recent literature, the Microsoft Quantum group's own prior work on InAs-Al hybrid devices passing the topological gap protocol (Ref. 60, Phys. Rev. B 2023) and interferometric single-shot parity measurements (Ref. 61, Nature 2025) establishes the immediate baseline. The present work demonstrates a three-order-of-magnitude improvement in parity lifetime (20 s versus 1–12 ms), which is substantial. However, the Kanne et al. Nature Nanotechnology 2021 work (Ref. 53) already demonstrated epitaxial Pb on InAs nanowires with 2e-periodic charging patterns consistent with hard induced gap. The key advance here is not the material system per se but the integration into a multi-tetron array geometry with demonstrable parity protection. Recent work by Song et al. (Nano Lett. 2025, Ref. 55) and Zhang et al. (Nano Lett. 2026, Ref. 56) on PbTe nanowires represents related but distinct materials approaches that do not yet demonstrate comparable parity lifetimes in qubit-relevant geometries.

The claimed parity lifetime of Z = 22 ś 1 s extracted from exponential fits to dwell time distributions (Fig. 6e) with N = 324 events raises questions about statistical power for distinguishing between competing decay models. The authors assume a homogeneous Poisson process, but non-exponential distributions could arise from time-varying quasiparticle densities or multiple competing poisoning mechanisms with different rates. With only 324 events aggregated across multiple measurements, distinguishing exponential from, say, stretched exponential or power-law tails would require careful model comparison that is not presented. The claim that "some instances reaching minute-scale" in the abstract is particularly concerning—this suggests significant variability that should be characterized systematically rather than highlighted anecdotally.

The comparison framework for establishing Pareto dominance over prior devices is implicitly presented but not rigorously constructed. The key performance metrics appear to be T (topological gap), EM (Majorana splitting), and Z (parity lifetime). For a Pareto improvement, the InAs-Pb devices should dominate on at least one metric without regression on others. The claimed improvements are T: 70 ţeV versus 30 ţeV (2.3Œ improvement), EM: < 1 ţeV (comparable or better), Z: 20 s versus 1–12 ms (>1000Œ improvement). However, the authors do not present direct measurements of X (X-parity lifetime) for the Pb-based devices, instead stating that "investigating these effects will be an important direction for future work." This is a significant gap since X scales as EMš, and improved Z without corresponding improvement in X would not constitute strict Pareto dominance for all qubit operations.

Questions for Authors:

1. Have you performed formal model selection (e.g., Akaike information criterion) comparing exponential versus non-exponential dwell time distributions for the parity switching data?
2. Can you provide even preliminary X measurements for the Pb-based devices to establish whether the claimed improvements extend to X-parity operations?
3. How do the fabrication yield and reproducibility of the InAs-Pb devices compare to the Al-based baseline, given that scalability is a central motivation?

Verdict: 6/10 — major-revisions

4.2.3 Voice 3 — Reviewer 3 (Empirical evidence)

The statistical treatment of the parity lifetime measurement (Fig. 6) requires more rigorous analysis than currently presented. The authors report Z = 22 ś 1 s based on exponential fitting of N = 324 dwell intervals aggregated across multiple measurements. The quoted uncertainty of ś1 s appears to represent only the fit uncertainty, not the full sampling variability. For a Poisson process with true rate , the maximum likelihood estimator of the mean dwell time has standard error /N 22/324 1.2 s, which is consistent with the reported uncertainty. However, the aggregation across "multiple measurements" (stated as 9 additional time traces beyond the example in Fig. 6d) and across a range of Vwp = 1.365 13 V ś 10 ţV introduces systematic variability that should be characterized. If the true parity lifetime varies with gate voltage even within this 20 ţV range, the aggregated distribution would exhibit over-dispersion relative to a homogeneous Poisson model.

The claim of "some instances reaching minute-scale" lifetime requires quantification. Examining Fig. 6e, the histogram extends to approximately 140 s with visible counts in the 100–140 s bins. However, for = 22 s, the probability of observing a dwell time exceeding 100 s is exp(100/22) 1%, predicting approximately 3 events in a sample of 324. The apparent presence of multiple events at 100+ s could indicate either a heavy tail inconsistent with exponential distribution or simply sampling variability. A quantile-quantile plot against the exponential distribution would clarify whether the tail behavior is anomalous.

The rf-based wire spectroscopy protocol (Section 3, Figs. 4–5) represents a systematic approach to identifying low-energy wire states, but the thresholding procedure introduces subjective elements that should be characterized through sensitivity analysis. The signal-to-noise threshold S/ > 2.5 and the requirement that at least 7/10 wire cutter configurations show above-threshold signal are reasonable but arbitrary. The authors acknowledge that "the exact boundary of the identified region is in general sensitive to the details of the thresholds used," but do not quantify this sensitivity. Varying these thresholds and reporting the resulting changes in the identified topological region would strengthen confidence that the conclusions are robust.

The manuscript lacks a systematic failure-modes section discussing devices or measurements that did not work as expected. While Section 3 mentions that "additional sign switches are interpreted to be the result of external quasiparticle poisoning events," this is the only acknowledgment of imperfect behavior. Questions arise naturally: What fraction of devices in the multi-tetron array achieved the reported performance? Were there systematic differences between the four tetrons (AA, AB, BA, BB) in the unit cell? Did all nanowires in the device exhibit comparably low EM, or is the top wire of the BA tetron a particularly favorable case? The reproducibility claims implicit in the scalability discussion would be better supported by explicit characterization of device-to-device and wire-to-wire variability within the array.

The data provenance and reproducibility infrastructure are not described. While the authors reference specific gate voltages (e.g., Vwp = 1.365 13 V) and magnetic fields (e.g., B = 3.8 T, Bz in Fig. 5), there is no mention of data availability statements, analysis code, or raw data repositories. For a result with significant implications for quantum computing scalability, the ability to independently verify numerical claims from archived data and analysis scripts would substantially enhance credibility. Even if full data release is not feasible for proprietary reasons, describing the audit trail from raw measurements to reported numbers would address concerns about replicability.

Questions for Authors:

1. Can you provide a quantile-quantile plot comparing the observed dwell time distribution against the fitted exponential, particularly characterizing the tail behavior that gives rise to "minute-scale" instances?
2. What is the wire-to-wire and device-to-device variability in EM and Z across the multi-tetron array, and what fraction of measured wires fall within the claimed performance specifications?
3. Is there a data availability statement or analysis code repository that would enable independent verification of the reported numerical results?

Verdict: 6/10 — major-revisions

4.2.4 Voice 4 — Devil's Advocate

This manuscript exhibits a concerning pattern of presenting aspirational claims with insufficient supporting evidence, while downplaying or omitting information that would complicate the triumphalist narrative. Let me be specific about the most serious problems that the other reviewers have been too generous in overlooking.

First, the three-order-of-magnitude improvement in parity lifetime is based on a single wire in a single device, measured over an unspecified total duration, with results aggregated across an unspecified number of measurement sessions potentially spanning days. The authors casually mention that "the measurements in

Fig. 4 and Fig. 5 were taken several days apart" with "some small shifts in gate voltages expected" (Ref. 85), revealing that device stability over relevant timescales is not actually demonstrated. If gate voltage drift of hundreds of microvolts occurs between measurement sessions, how can we be confident that the 20-second parity lifetime would be maintained over the operational timescales required for quantum computation? The entire premise of using parity lifetime as a qubit metric assumes stable operating conditions, yet stability is neither demonstrated nor even claimed.

Second, the claimed topological gap of T 70 ţeV (top quintile) is presented as a definitive improvement, but this number comes from the Topological Gap Protocol applied to test device structures (Fig. 3c), not to the actual multi-tetron array (Fig. 1) in which the parity measurements were performed. The TGP measurement was done on a "3 ţm-long nanowire test structure" according to the figure caption, whereas the tetron nanowires are 3.5 ţm long with different geometry including the narrow backbone junction. The authors provide no evidence that the TGP results transfer to the actual qubit geometry. This is not a minor caveat—the relationship between test structure performance and integrated device performance is the central question for any claim about scalability.

Third, the rf-based wire spectroscopy method, presented as enabling "scalable tuning," has only been demonstrated on one wire of one tetron. The multi-tetron array shown in Fig. 1 contains eight nanowires (two per tetron, four tetrons), yet all the detailed measurements (Figs. 4–6) focus exclusively on the top wire of the BA tetron. Where is the data from the other seven wires? If this is truly a "prototype unit cell for multi-tetron devices" suitable for "scaling to larger qubit arrays," why is there no demonstration of parallel tuning and characterization? The closest the authors come is mentioning that "all of the QDs have been designed to have plunger lever arms in the range 0.4–0.45 meV/mV," which is a design claim, not an experimental demonstration.

Fourth, the attribution of improved parity lifetime to the Pb-Al material difference is correlational, not causal. The authors changed multiple variables simultaneously: the superconductor (Al to Pb), the substrate (implied InP to GaSb), the quantum well composition (adding InAsSb), and presumably numerous fabrication details. While they argue that the larger Pb gap and stronger electron-phonon coupling are responsible for the improvement, they provide no controlled experiment varying only the superconductor while holding other factors constant. The device measured here is not a simple material substitution—it is a complete redesign. Alternative explanations, such as reduced defect density from the GaSb substrate or reduced quasiparticle injection from improved shielding, cannot be excluded.

The manuscript's discussion of implications for fault-tolerant quantum computing is premature to the point of being misleading. The abstract claims that "non-equilibrium quasiparticles no longer limit qubit operations in our devices," but no actual qubit operations are demonstrated. There is no coherent manipulation, no gate fidelity measurement, no demonstration of even a single logical operation. The parity measurement is a necessary but not sufficient component of a functioning qubit. By the same logic, I could claim that a very stable piece of iron "no longer limits" magnetic memory operations because it maintains its magnetization for years—but this says nothing about whether it can function as a memory element in an actual computing system.

Recommendation: major-revisions

The paper presents genuinely interesting results on material development and measurement techniques for Majorana-based devices, but the claims about qubit performance and scalability are substantially oversold relative to the evidence presented. Major revision is required to either scale back the claims to match the evidence or to provide substantially more data supporting the scalability and reproducibility assertions.

4.2.5 Voice 5 — Editor-in-Chief synthesis

Having considered all four reviews, I find a manuscript that presents legitimate scientific advances alongside claims that substantially exceed the evidentiary support. The disagreement between reviewers reflects genuine tension in the paper between high-quality materials characterization and overreaching implications for quantum computing.

Reviewer 1 finds the physics framework sound but identifies gaps in the quantitative justification, particularly regarding induced gap simulations, spin-orbit coupling extraction methodology, and the connection between the minimal Hamiltonian model and the claim of ţeV-resolution EM extraction. These are addressable through additional analysis and clearer presentation of methodology, consistent with minor revisions. Reviewer 2 raises more serious concerns about the novelty claim structure, correctly identifying that the absence of X measurements for Pb-based devices undermines the Pareto dominance argument for complete qubit performance. This is a substantive gap that requires additional experimental data. Reviewer 3's concerns about statistical rigor, threshold sensitivity in the tuning protocol, and absence of failure-mode documentation are well-founded and align with best practices for high-impact experimental claims.

The Devil's Advocate raises the most damaging critique: the disconnect between the claimed "scalable multi-tetron array" architecture and the actual measurement scope, which focuses exclusively on one wire of one tetron. This criticism is valid and substantially undermines the paper's central narrative about demonstrating a pathway to fault-tolerant quantum computing. The observation that TGP measurements come from test structures rather than the actual qubit device is particularly concerning, as it breaks the evidentiary chain linking material improvements to qubit performance. However, I do not agree with the implicit suggestion that the entire line of reasoning is invalid—the improvements in parity lifetime are real and significant, even if the scalability claims are premature.

Reconciling these perspectives, I find that the manuscript would be appropriate for PRX Quantum after major revisions. The core scientific contribution—demonstration of dramatically improved parity lifetimes in Pb-based devices combined with a scalable rf-based tuning methodology—is significant and timely. However, the presentation must be substantially revised to align claims with evidence. The scalability narrative should be tempered to acknowledge that multi-wire, multi-tetron characterization remains future work. The failure-mode and reproducibility documentation needs expansion. And the statistical treatment of parity lifetime measurements requires the rigor appropriate for a flagship result.

Must-fix items before resubmission, ordered by severity:

1. Provide characterization data from at least one additional wire in the multi-tetron array (preferably from a different tetron than BA) demonstrating comparable EM and parity lifetime, or revise all claims about scalability and array-level performance to explicitly acknowledge single-wire scope.
2. Include either preliminary X measurements for the Pb-based devices or remove all claims about implications for X-parity operations and general qubit performance, explicitly scoping the results to Z-parity measurements only.
3. Present quantitative model comparison (AIC or similar) between exponential and alternative dwell time distributions, including explicit treatment of the "minute-scale instances" tail behavior and characterization of measurement-to-measurement variability in extracted Z.
4. Perform and report sensitivity analysis on the rf tuning protocol thresholds, quantifying how the identified topological region changes with S/ threshold and cutter configuration requirements.
5. Either present TGP measurements on the actual tetron device in which parity measurements were performed, or clearly separate test structure characterization from device-level performance with explicit discussion of the expected transfer of properties.
6. Add a data availability statement specifying the accessibility of raw measurement data and analysis code, consistent with PRX Quantum policies on reproducibility.

4.2.6 Vote table

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

4.3 majorana2 — Logical-fallacy report

Model: `claude-opus-4-5-20251101` · 3,112 output tokens · 62.7 s

Fallacy: cherry-picked-baseline **Severity:** high **Location:** Section 1 (Introduction), paragraph 2 and throughout **Evidence:** “In earlier Al-based devices, we observed typical top quintile gaps of T 30 ţeV [59–61]. In contrast, in the InAs–Pb devices studied here we observe a top quintile gap of T 70 ţeV.” **Why it’s the fallacy:** The manuscript exclusively compares its Pb-based devices against the authors’ own earlier Al-based devices from their own lab (refs 59-61). While refs 53-56 mention other groups’ Pb-based nanowire work, the paper dismisses this prior art with the vague claim that they have “gone beyond previous work which has incorporated larger gap superconductors into nanowire devices [53–56]” without providing any quantitative comparison of topological gaps, parity lifetimes, or other metrics against these published baselines. This selective comparison against only weaker internal baselines while ignoring potentially stronger external results constitutes cherry-picking. **Suggested fix:** Add a table comparing quantitative metrics (T, parity lifetime, EM) against the specific results from refs 53-56 and any other relevant Pb-based or high-gap superconductor nanowire studies. If those studies didn’t report comparable metrics, state this explicitly.

Fallacy: hasty-generalization **Severity:** medium **Location:** Section 4 (Interferometric Parity Readout), paragraph on parity lifetime **Evidence:** “To quantify the lifetime we classify the data by fitting a Gaussian mixture model… By aggregating multiple measurements, we observe a total of N = 324 dwell intervals. The data are consistent with a single exponential distribution… We extract a characteristic parity lifetime and corresponding fit uncertainty of Z = 22 ś 1 s” **Why it’s the fallacy:** The 20+ second parity lifetime claim—which is the paper’s headline result—is derived from measurements on a single nanowire of a single tetron in a single device. The abstract and conclusions generalize this to “InAs–Pb tetron devices” broadly, but there is no demonstration that this result is reproducible across multiple devices, multiple tetrons, or even multiple wires within the same device. The measurement also appears to be taken at a specific optimized operating point (Vwp = 1.365 13 V). **Suggested fix:** Either present parity lifetime statistics from multiple devices/tetrons/wires, or qualify claims to state “In a single nanowire of one tetron, we measured… ” and explicitly note this as a demonstration requiring broader validation.

Fallacy: conflated-regimes **Severity:** medium **Location:** Section 5 (Discussion and Outlook), final paragraphs **Evidence:** “The multi-tetron array presented here functions as a modular ‘unit cell’ for a larger architecture; it can be tiled into much larger qubit arrays (e.g., a 12-qubit array) without altering the underlying control or readout approach. The strong parity protection observed in our tetron prototype suggests that, even as the system scales up, each qubit will remain well isolated from non-equilibrium quasiparticles” **Why it’s the fallacy:** The paper extrapolates from a single tetron measurement (one wire, one parity lifetime measurement) to claims about how a 12-qubit array would behave. This conflates the regime of a single isolated device with the regime of scaled arrays where crosstalk, thermal load from additional control lines, and other scaling-specific noise sources may dominate. No evidence is provided that scaling preserves these properties. **Suggested fix:** Reframe as: “Demonstrating similar performance in scaled arrays remains an important open challenge. Potential scaling concerns include [list specific concerns] which will require investigation in multi-qubit devices.”

Fallacy: active-space-handwave **Severity:** medium **Location:** Section 5 (Discussion and Outlook) **Evidence:** “Looking ahead, the values of EM achieved here are expected to have beneficial implications for Pauli-X measurements, whose characteristic switching time X scales as EMš. Deep in the topological regime, EM T exp(L/T)… Increasing the NW length L exponentially suppresses EM and thus dramatically extends X, which suggests that X could be more than an order of magnitude longer than in previous devices.” **Why it’s the fallacy:** The paper claims that X “could be more than an order of magnitude longer” based on theoretical scaling arguments, but does not actually perform X-parity measurements or demonstrate this. The claim of generalizing to X measurements is made without running the experiment, which constitutes an active-space handwave—claiming a result extends to a regime not actually tested. **Suggested fix:** Remove the speculative quantitative claim (“more than an order of magnitude”) or qualify it as: “Based on theoretical scaling relations, we predict X improvements, but experimental validation of X-parity lifetimes in Pb-based devices remains for future work.”

Fallacy: ad-hoc-precision-floor **Severity:** medium **Location:** Section 3 (RF-Based Wire Spectroscopy), discussion of EM resolution **Evidence:** “we expect the resolution of this analysis to be limited to 1 ţeV… Notably, the resolution of 1 ţeV significantly exceeds that of conductance measurements which can resolve EM of about the half-width-half-max of a temperature broadened conductance peak EM 1.76kB T 7.6 ţeV for T = 50 mK.” **Why it’s the fallacy:** The paper claims EM values “below our 1 ţeV resolution” as evidence of topological behavior, but this resolution floor is derived from DAC resolution and lever arm estimates, not from demonstrated noise floors or calibration standards. The comparison to “7.6 ţeV” for conductance measurements also involves mixing different measurement modalities and conditions without rigorous cross-calibration. Claiming sub-resolution EM values as a positive result approaches the ad-hoc-precision-floor fallacy. **Suggested fix:** Provide explicit calibration of the 1 ţeV resolution claim with known energy scales, or reframe as: “We observe EM consistent with zero within our measurement resolution of approximately 1 ţeV, though we cannot exclude finite EM below this threshold.”

Fallacy: appeal-to-authority **Severity:** medium **Location:** Section 1 (Introduction), opening paragraph **Evidence:** “Recently, we presented a roadmap [1] to fault-tolerant quantum computation using topological qubits [2–4] built around Majorana zero modes (MZMs) in superconductor-semiconductor hybrid devices [5–9]. Our roadmap draws on concepts explored in Refs. 10–39.” **Why it’s the fallacy:** The paper opens by citing 39 references in the first paragraph, with 30 of them (refs 10-39) bundled into a single “draws on concepts” citation. This mass citation serves more as an appeal to the authority and weight of the cited literature than as specific technical justification. The roadmap ref [1] is the authors’ own prior publication, creating a self-referential authority structure. **Suggested fix:** Cite specific concepts from specific papers where they are used technically, rather than bundling 30 references as general background authority. If a roadmap citation is needed, briefly state which specific elements are being implemented.

Fallacy: equivocation **Severity:** medium **Location:** Throughout, particularly Abstract and Section 4 **Evidence:** Abstract: “parity lifetime of 20 s with some instances reaching minute-scale”; Section 4: “intrinsic Z-parity lifetime of the device which far exceeds the measurement time” **Why it’s the fallacy:** The paper uses “parity lifetime” to refer to multiple related but distinct quantities: (1) Z-parity lifetime in an interference measurement, (2) quasiparticle poisoning time, and (3) intrinsic parity switching time. The headline “20 second parity lifetime” refers specifically to Z-parity in one measurement configuration, but the discussion sometimes conflates this with general qubit lifetime or poisoning immunity. The phrase “some instances reaching minute-scale” is particularly vague about what operational definition of parity lifetime is being used. **Suggested fix:** Define “Z-parity lifetime” precisely at first use and use consistent terminology throughout. Clarify whether “minute-scale” instances represent statistical fluctuations in the exponential distribution or a different measurement condition.

4.4 majorana2 — Argument-structure audit

Model: `claude-haiku-4-5-20251001` · 3,083 output tokens · 73.5 s

ARGUMENT STRUCTURE REPORT

4.4.1 Executive summary

Overall verdict: PUBLISH-READY **Claim–proof gap:** NONE — The paper claims that replacing Al with Pb in tetron devices yields larger topological gaps and longer parity lifetimes, and demonstrates exactly this through material characterization, TGP analysis, and time-resolved parity measurements. **CME balance:** BALANCED — C:25% / M:30% / E:45% **Narrative debts:** 5 total (0 load-bearing) **Sequencing:** EVIDENTIAL

4.4.2 Argument map

P1 (premise, ğ1): Topological protection gives error rates exponentially suppressed as the topological ga

P2 (premise, ğ1): The parent superconducting gap governs the topological gap in superconductor-semiconductor hybrids [Refs. 4052].

P3 (premise, ğ2): Pb has a parent gap Pb 1.3 meV vs Al 300 ţeV.

P4 (premise, ğ2): The GaSb substrate enables lattice-matched growth and larger spin-orbit coupling (~ 12

I1 (intermediate, from P2+P3+P4): InAsPb devices should exhibit larger topological gaps than Al-based devices.

E1 (evidence, ğ23): TGP analysis shows top-quintile T ~ 70 ţeV in Pb devices vs ~30 ţeV in Al devices.

E2 (evidence, ğ3): Localization lengths exceed 1 ţm, confirming low disorder.

E3 (evidence, ğ3): Zero-bias peaks persist over >0.5 T field range at both wire ends.

I2 (intermediate, from I1+E1+E2+E3): The Pb-based platform achieves a robust topological phase with enlar

P5 (premise, ğ1): Non-equilibrium quasiparticles limit parity lifetime independently of T.

P6 (premise, ğ4): Pb’s higher gap suppresses Cooper pair breaking and enhances recombination.

E4 (evidence, ğ4): Measured Z-parity lifetime Z = 22 ś 1 s via interferometric readout with h/2e periodic

E5 (evidence, ğ4): Prior Al-based devices showed Z ~ 112 ms [Refs. 59, 61].

I3 (intermediate, from P5+P6+E4+E5): Parity lifetime improved by >3 orders of magnitude due to Pb substit

E6 (evidence, ğ3): rf spectroscopy shows EM < 1 ţeV across extended parameter regimes.

I4 (intermediate, from E6): Majorana hybridization is suppressed below measurement resolution.

C (conclusion, from I2+I3+I4): Increasing the excitation gap via Pb substitution directly translates to i

Unsupported leaps: None identified. Each intermediate claim follows from stated premises plus presented evidence.

Unstated premises: 1. The parity lifetime measurement faithfully reflects quasiparticle poisoning rate rather than alternative relaxation mechanisms (partially addressed in ğ4 discussion of Poisson statistics). 2. The single measured tetron is representative of the array; no cross-device statistics are shown for parity lifetime. 3. Floating tetrons will have comparable or better parity lifetimes (acknowledged as expectation, not demonstrated).

4.4.3 Section A: Controlling idea

(a) STATED CLAIM

“Here, we experimentally validate this principle in an InAs–Pb tetron device via interferometric single-shot parity measurements. By replacing aluminum with the higher-gap superconductor lead in our superconductor-semiconductor hybrid devices, we have improved the robustness of our topological phase.”

The strongest single sentence appears in ğ5:

> “Our results confirm a central premise of topological quantum computing: increasing the excitation gap dramatically reduces error mechanisms and improves qubit performance.”

(b) DEMONSTRATED CONCLUSION

An InAs–Pb tetron device exhibits a measured topological gap of ~70 ţeV (vs ~30 ţeV in Al), Majorana splitting EM < 1 ţeV, and Z-parity lifetime of ~22 s (vs ~1–12 ms in Al), directly demonstrating that substituting a higher-gap superconductor yields quantitative performance improvements across multiple metrics.

(c) CLAIM–PROOF GAP

NONE — The abstract and introduction claim that increasing the gap improves device performance, and the evidence architecture establishes exactly this through: (i) material characterization showing higher induced gap, (ii) TGP analysis showing larger topological gap, (iii) rf spectroscopy showing low EM, and (iv) time-resolved parity measurement showing ~22 s lifetime. The comparison to prior Al-based devices (Refs. 59, 61) anchors the improvement claim. The paper does not overclaim universality or fault-tolerance; it states these are "implications" and "expectations" for future work.

4.4.4 Section B: CME proportionality

CLAIM (C): STRONG — ~25%

Claims are made in ğ1 (introduction) and ğ5 (discussion): improved topological gap, suppressed EM, orders-of-magnitude improvement in parity lifetime, scalability implications. Claims are appropriately scoped to what is demonstrated and clearly distinguish demonstrated results from extrapolated expectations (e.g., "we expect that parity lifetimes in floating tetrons could be even longer").

MECHANISM (M): STRONG — ~30%

ğ1 provides the theoretical framework (topological protection scales with gap, quasiparticle poisoning independent of T). ğ2 explains why Pb + GaSb substrate should yield improvements (higher parent gap, lattice-matched growth, enhanced spin-orbit). ğ3 (rf spectroscopy) derives the single-state model Hamiltonian (Eq. 1) and quantum capacitance response (Eq. 2). ğ4 explains the physical origin of parity lifetime enhancement (harder to break Cooper pairs in Pb, faster recombination). The mechanism story is complete and quantitative.

EVIDENCE (E): STRONG — ~45%

The paper is evidence-rich: - Material characterization: Fig. 2 (induced gap 400–570 ţeV, mobility >350,000 cmš/Vs, spin-orbit 12ś2 meV nm) - TGP analysis: Fig. 3 (T ~70 ţeV, topological region >1.1 mVůT, zero-bias peaks over >0.5 T) - rf spectroscopy: Figs. 4–5 (EM extraction with ~1 ţeV resolution, correlation maps) - Parity measurement: Fig. 6 (h/2e periodicity, Z = 22ś1 s from exponential fit to N=324 dwell intervals)

Verdict: BALANCED

C:25% / M:30% / E:45%. No dimension dominates by more than ~20 points; evidence is the largest component but appropriately so for an experimental paper.

4.4.5 Section C: Narrative-debt register

Promise type	Promise (location)	Status	Load-bearing?
EVIDENCE PROMISE	"we experimentally validate this principle" (abstract)	FULFILLED	N/A
EVIDENCE PROMISE	"we have developed an rf measurement technique that resolves low-energy wire-end states and directly measures their energy splitting with ţeV precision" (abstract)	FULFILLED — ğ3 demonstrates resolution ~1 ţeV	No
EVIDENCE PROMISE	"We employ this technique to bring up a device in a multi-tetron array" (abstract)	FULFILLED — ğ3–4 operate on the BA tetron in the array	No
SCOPE PROMISE	"Further time-resolved measurements reveal a characteristic parity switching time of ~20 s with some instances reaching minute-scale" (abstract)	PARTIAL — Z = 22ś1 s is shown; "minute-scale" instances are not explicitly displayed in Fig. 6(e) distribution	Cosmetic

Promise type	Promise (location)	Status	Load-bearing?
RHETORICAL QUESTION	“we discuss potential implications for the fidelity of Pauli measurements” (abstract)	PARTIAL — ğ5 discusses qualitatively but does not quantify fidelity	Cosmetic

Total narrative debts: 5 (0 load-bearing)

The “minute-scale” claim is supported by the tail of the exponential distribution but no individual trace is highlighted; this is minor. The fidelity discussion in ğ5 is qualitative, but the abstract uses “discuss” rather than “quantify,” so no strong expectation is violated.

4.4.6 Section D: Sequencing diagnosis

Verdict: EVIDENTIAL

The paper follows a classic escalation structure: 1. ğ1 (Introduction): Establishes premises (topological protection, roadmap, error mechanisms) 2. ğ2 (Device design and materials): Describes platform and justifies why it should work 3. ğ3 (rf spectroscopy): Introduces new characterization method, demonstrates low EM 4. ğ4 (Parity readout): Presents the headline result (Z ~ 20 s) 5. ğ5 (Discussion): Synthesizes and extrapolates

The strongest result (parity lifetime) appears late, after the evidential scaffolding is complete. The abstract front-loads the result, which is standard for physics journals and does not constitute headline-first body organization.

Proposed resequencing moves: None required. The current order supports evidential escalation effectively. One optional refinement:

Current position	Optimal position	Benefit
ğ5 paragraph on X scaling (“the values of EM achieved here are expected to have beneficial implications. . . ”)	Could appear earlier in ğ4 as a framing paragraph	Would connect EM measurements (ğ3) to parity measurements (ğ4) more tightly before presenting data

This is minor and not essential.

4.4.7 Section E: Structural gaps

Missing analysis	Where it should appear	CME dimension strengthened
Cross-device statistics for parity lifetime	ğ4 or supplementary	Evidence — currently N=1 tetron; showing variance across array would strengthen scalability claim
Error bars on T values	ğ3 (Fig. 3c)	Evidence — “top quintile” is stated but distribution not shown
Direct comparison table: Al vs Pb metrics	ğ5 or ğ2	Claim — would make the “3 orders of magnitude” comparison more immediate
Noise model for CQ telegraph noise	ğ4	Mechanism — would formalize the Poisson assumption and bound systematic errors
X measurement or projection	ğ5	Evidence — discussed qualitatively but not measured; would discharge the Pauli-X implication promise

None of these gaps are critical; the paper is complete for its stated scope.

4.4.8 Summary diagnosis

The paper claims that substituting lead for aluminum in a tetron device yields dramatically improved topological protection, and demonstrates exactly this: a measured topological gap ~2Œ larger, Majorana splitting below 1 ţeV, and parity lifetime >1000Œ longer than prior Al-based devices. The argument is evidentially complete, with strong material characterization (TEM, Hall bars, SdH), protocol development (rf spectroscopy), and direct measurement (h/2e-periodic parity readout with Z = 22ś1 s). The CME balance is appropriate for an experimental paper; no dimension is absent or critically thin. Narrative debts are cosmetic. Sequencing is evidential. The paper is what it claims to be: an experimental validation that increasing the superconducting gap improves topological qubit performance. No restructuring is required; minor additions (cross-device statistics, explicit comparison table) would strengthen but are not essential.

4.5 majorana2 — Numeric-claim registry (deterministic)

Numeric-Claim Registry Audit Paper: `majorana2.pdf` — 2 sections indexed, 0 distinct ground-truth values (from none).

Not applicable: no Results/Experiments/Tables section was identified, so there is no in-paper ground truth to gate against. This is expected for letter-style papers without named section headings or for lossy PDF extraction. Run against the LaTeX/markdown source, or pass –strict-on with this paper's actual section names.

Clean. Every numeric in the non-strict sections traces back to a value present in Results / Experiments / Tables.

4.6 majorana2 — Disclosure audit

Model: `claude-opus-4-5-20251101` · 1,781 output tokens · 33.7 s

DISCLOSURE AUDIT REPORT

4.6.1 Compliance status by category

Code	Category	Status	Evidence (section or "absent")	Required for venue?
A1	Funding statement	ABSENT	absent	Yes
A2	Competing interests / COI	ABSENT	absent	Yes
A3	Author contributions	ABSENT	absent	Yes
A4	Data availability	ABSENT	absent	Yes
A5	Code availability	ABSENT	absent	Yes
A6	Ethics / IRB approval	NOT-APPLICABLE	Pure device/experimental physics paper with no human/animal subjects	No
A7	Preprint / prior-publication status	PRESENT-COMPLETE	"arXiv:2606.03884v1 [cond-mat.mes-hall] 2 Jun 2026" on page 1	Yes
A8	Materials availability	ABSENT	absent	Yes
B1	AI-assisted text drafting	ABSENT	absent	Yes (if applicable)
B2	AI-generated images/figures	ABSENT	absent	Yes (if applicable)
B3	AI-assisted data analysis or coding	ABSENT	absent	Yes (if applicable)
B4	AI-assisted pre-submission review	ABSENT	absent	Yes (if applicable)
C1	Prior-publication warranty	ABSENT	absent	Yes

Code	Category	Status	Evidence (section or "absent")	Required for venue?
C2	Government / institutional / export-control approval	ABSENT	absent	Yes (quantum hardware work)
C3	Third-party rights	ABSENT	absent	Yes (if applicable)
C4	Free-online-version conflict	PRESENT-INCOMPLETE	arXiv preprint exists; PRX Quantum permits arXiv but manuscript lacks explicit statement of compliance	Yes

4.6.2 Missing disclosures and warranty gaps

Submission-blocking

- **A1 (Funding statement)**: No funding sources or grant numbers disclosed. Microsoft Quantum authorship implies corporate funding but explicit statement required.
- **A2 (Competing interests)**: No COI declaration. All authors affiliated with Microsoft Quantum, which has commercial interest in quantum computing; financial/employment conflicts must be declared.
- **A3 (Author contributions)**: Group authorship "Microsoft Quantum" with footnote listing 150+ contributors but no CRediT-style role assignments.
- **A4 (Data availability)**: No statement regarding experimental data, processed datasets, or access procedures.
- **A5 (Code availability)**: References to simulations and analysis code (e.g., TGP protocol, fitting procedures) but no repository, license, or availability statement.
- **C1 (Prior-publication warranty)**: No explicit statement that manuscript is not under consideration elsewhere.

Revise-before-acceptance

- **A8 (Materials availability)**: Device fabrication involves proprietary InAs–Pb heterostructures, GaSb substrates, and specific epitaxial growth. No MTA information or availability statement provided.
- **C2 (Government/institutional/export-control approval)**: Quantum computing hardware may be subject to export controls; no pre-publication clearance statement.
- **B1–B4 (AI disclosures)**: PRX Quantum increasingly requires disclosure of AI assistance. Authors must confirm whether AI tools were used and provide explicit negative statement if not.

Editorial-discretion

- **C3 (Third-party rights)**: All figures appear original. Confirm no reuse requiring permission.
- **C4 (Free-online-version conflict)**: arXiv posting compatible with PRX Quantum policy but explicit acknowledgment recommended.

4.6.3 Submission-ready checklist

Code	Status	Fix Required
A1	FAIL	Add funding statement with grant numbers or "This work was funded internally by Microsoft Corporation with no external grant support."
A2	FAIL	Add competing interests declaration: all authors are employees of Microsoft Corporation which is developing commercial quantum computing technology.
A3	FAIL	Add CRediT-style author contributions for the 150+ listed contributors or designate writing committee with specific roles.
A4	FAIL	Add data availability statement specifying repository/access or "Data available from corresponding author upon reasonable request."

Code	Status	Fix Required
A5	FAIL	Add code availability statement with repository URL and license, or explicit “Code available upon reasonable request” with justification.
A6	PASS	Not applicable for device physics paper.
A7	PASS	arXiv ID present: 2606.03884v1.
A8	FAIL	Add materials availability statement addressing InAs–Pb heterostructure availability, fabrication protocols, and MTA requirements.
B1	FAIL	Add explicit statement: “AI tools were [not] used for text drafting” with model/scope if applicable.
B2	FAIL	Add explicit statement: “All figures were created without AI assistance” or disclose tools used.
B3	FAIL	Add explicit statement regarding AI use in data analysis/coding or confirm none used.
B4	FAIL	Add explicit statement regarding AI-assisted review or confirm none used.
C1	FAIL	Add warranty: “This manuscript has not been published previously and is not under consideration at another journal.”
C2	FAIL	Add statement confirming institutional/export-control pre-publication approval for quantum hardware disclosure.
C3	PASS	Requires author confirmation that all figures are original; no obvious third-party content identified.
C4	PASS	arXiv preprint policy compatible with PRX Quantum; recommend adding explicit statement.

4.7 majorana2 — Anchored revision plan

Model: `claude-haiku-4-5-20251001` · 2,676 output tokens · 56.1 s

Anchored Revision Roadmap

4.7.1 1. Provide characterization data from at least one additional wire in the multi-tetron array (preferably from a different tetron than BA) demonstrating comparable EM and parity lifetime, or revise all claims about scalability and array-level performance to explicitly acknowledge single-wire scope.

Severity: 5 **Effort**: multi-day **Judges**: Reviewer 2, Reviewer 3, Devil’s Advocate, Editor-in-Chief

Source paragraph(s): ű003, ű008, ű024

Quoted problem prose (verbatim from the manuscript, <=2 sentences): > “We employ this technique to bring up a device in a multi-tetron array and perform parity measurements of one of the tetron’s hybrid nanowires (NWs).”

Judge evidence (one bullet per judge that diagnosed this; quote the judge verbatim, <=1 sentence each): - Reviewer 3: “What is the wire-to-wire and device-to-device variability in EM and Z across the multi-tetron array, and what fraction of measured wires fall within the claimed performance specifications?” - Devil’s Advocate: “The multi-tetron array shown in Fig. 1 contains eight nanowires (two per tetron, four tetrons), yet all the detailed measurements (Figs. 4–6) focus exclusively on the top wire of the BA tetron.” - Editor-in-Chief: “Provide characterization data from at least one additional wire in the multi-tetron array (preferably from a different tetron than BA) demonstrating comparable EM and parity lifetime, or revise all claims about scalability and array-level performance to explicitly acknowledge single-wire scope.”

Proposed edit: Either add a supplementary figure presenting EM extraction and parity lifetime data from at least one additional wire (e.g., the bottom wire of BA or any wire from AA, AB, or BB tetron), or revise ű003 and ű024 to read: “We employ this technique to bring up a device in a multi-tetron array and perform parity measurements on *one nanowire of one tetron*; characterization of array-wide performance remains for future work.”

Why this works: Explicitly scoping the measurement to a single wire eliminates the mismatch between the claimed “scalable multi-tetron array” framing and the single-wire evidence actually presented.

4.7.2 2. Include either preliminary X measurements for the Pb-based devices or remove all claims about implications for X-parity operations and general qubit performance, explicitly scoping the results to Z-parity measurements only.

Severity: 4 **Effort**: 2 h (if removing claims) or multi-day (if adding measurements) **Judges**: Reviewer 2, Editor-in-Chief

Source paragraph(s): ű024

Quoted problem prose (verbatim from the manuscript, <=2 sentences): > “Looking ahead, the values of EM achieved here are expected to have beneficial implications for Pauli-X measurements, whose characteristic switching time X scales as EMš. … Investigating these effects will be an important direction for future work.”

Judge evidence (one bullet per judge that diagnosed this; quote the judge verbatim, <=1 sentence each): - Reviewer 2: “Can you provide even preliminary X measurements for the Pb-based devices to establish whether the claimed improvements extend to X-parity operations?” - Editor-in-Chief: “Include either preliminary X measurements for the Pb-based devices or remove all claims about implications for X-parity operations and general qubit performance, explicitly scoping the results to Z-parity measurements only.”

Proposed edit: If X data are unavailable, revise ű024 to delete the sentence “which suggests that X could be more than an order of magnitude longer than in previous devices” and replace with: “Experimental characterization of X in Pb-based devices remains an open direction; the results presented here are limited to Z-parity measurements.”

Why this works: Removes speculative quantitative claims about unmeasured quantities, aligning claims with the presented evidence.

4.7.3 3. Present quantitative model comparison (AIC or similar) between exponential and alternative dwell time distributions, including explicit treatment of the “minute-scale instances” tail behavior and characterization of measurement-to-measurement variability in extracted Z.

Severity: 4 **Effort**: 2 h **Judges**: Reviewer 2, Reviewer 3, Editor-in-Chief

Source paragraph(s): ű003, ű022

Quoted problem prose (verbatim from the manuscript, <=2 sentences): > “Further time-resolved measurements reveal a characteristic parity switching time of 20 s with some instances reaching minute-scale.”

Judge evidence (one bullet per judge that diagnosed this; quote the judge verbatim, <=1 sentence each): - Reviewer 2: “Have you performed formal model selection (e.g., Akaike information criterion) comparing exponential versus non-exponential dwell time distributions for the parity switching data?” - Reviewer 3: “Can you provide a quantile-quantile plot comparing the observed dwell time distribution against the fitted exponential, particularly characterizing the tail behavior that gives rise to ‘minute-scale’ instances?” - Editor-in-Chief: “Present quantitative model comparison (AIC or similar) between exponential and alternative dwell time distributions, including explicit treatment of the ‘minute-scale instances’ tail behavior and characterization of measurement-to-measurement variability in extracted Z.”

Proposed edit: Add to ű022 or a supplementary section: (1) a Q-Q plot of observed dwell times vs. the exponential(=22 s) quantiles, (2) AIC values comparing exponential, stretched-exponential, and power-law fits, and (3) a table showing Z extracted from each of the 9 individual time traces with standard deviation to characterize run-to-run variability.

Why this works: Demonstrates that the exponential assumption is justified (or identifies systematic deviations), and quantifies the “minute-scale” tail relative to statistical expectation.

4.7.4 4. Perform and report sensitivity analysis on the rf tuning protocol thresholds, quantifying how the identified topological region changes with S/ threshold and cutter configuration requirements.

Severity: 3 **Effort**: 2 h **Judges**: Reviewer 3, Editor-in-Chief

Source paragraph(s): ű020

Quoted problem prose (verbatim from the manuscript, <=2 sentences): > “To identify regions associated with stable low-energy states, we evaluate, for each point in Vwp and Bz, the fraction of wire-cutter configurations for which the rf response exceeds a signal-to-noise ratio S/ > 2.5.”

Judge evidence (one bullet per judge that diagnosed this; quote the judge verbatim, <=1 sentence each): - Reviewer 3: “The thresholding procedure introduces subjective elements that should be characterized through sensitivity analysis.” - Editor-in-Chief: “Perform and report sensitivity analysis on the rf tuning protocol thresholds, quantifying how the identified topological region changes with S/ threshold and cutter configuration requirements.”

Proposed edit: Add a supplementary figure showing the identified region of interest (analogous to Fig. 5c) for S/ thresholds of 2.0, 2.5, and 3.0, and for requiring 5/10, 7/10, and 9/10 cutter configurations. Include a brief statement in ű020: “Varying the S/ threshold from 2.0 to 3.0 shifts the region boundary by approximately X mV in Vwp; qualitative features are preserved.”

Why this works: Demonstrates robustness (or quantifies sensitivity) of the tuning protocol to threshold choices, addressing concerns about subjectivity.

4.7.5 5. Either present TGP measurements on the actual tetron device in which parity measurements were performed, or clearly separate test structure characterization from device-level performance with explicit discussion of the expected transfer of properties.

Severity: 4 **Effort**: 30 min (if clarifying scope) or multi-day (if adding TGP data) **Judges**: Devil’s Advocate, Editor-in-Chief

Source paragraph(s): ű012

Quoted problem prose (verbatim from the manuscript, <=2 sentences): > “A representative phase diagram obtained from a NW test device is shown in Fig. 3(c), which was obtained by applying the Topological Gap Protocol (TGP) [60].”

Judge evidence (one bullet per judge that diagnosed this; quote the judge verbatim, <=1 sentence each): - Devil’s Advocate: “The TGP measurement was done on a ‘3 ţm-long nanowire test structure’ according to the figure caption, whereas the tetron nanowires are 3.5 ţm long with different geometry including the narrow backbone junction.” - Editor-in-Chief: “Either present TGP measurements on the actual tetron device in which parity measurements were performed, or clearly separate test structure characterization from device-level performance with explicit discussion of the expected transfer of properties.”

Proposed edit: Add to ű012 or the Fig. 3 caption: “The TGP data shown here were obtained from a standalone 3 ţm test nanowire; TGP measurements on the integrated tetron device were not performed due to [reason, e.g., lack of transport terminals]. We expect qualitatively similar behavior in the 3.5 ţm tetron wires based on [shared material stack / comparable localization lengths], though quantitative T values may differ.”

Why this works: Breaks the implicit claim that Fig. 3 directly characterizes the device in Figs. 4–6, providing transparency about the evidentiary gap.

4.7.6 6. Add a data availability statement specifying the accessibility of raw measurement data and analysis code, consistent with PRX Quantum policies on reproducibility.

Severity: 3 **Effort**: 5 min **Judges**: Reviewer 3, Editor-in-Chief

Source paragraph(s): ű? (structural finding—no data availability section exists)

Quoted problem prose (verbatim from the manuscript, <=2 sentences): > (no verbatim anchor — structural finding)

Judge evidence (one bullet per judge that diagnosed this; quote the judge verbatim, <=1 sentence each): - Reviewer 3: “Is there a data availability statement or analysis code repository that would enable independent verification of the reported numerical results?” - Editor-in-Chief: “Add a data availability statement specifying the accessibility of raw measurement data and analysis code, consistent with PRX Quantum policies on reproducibility.”

Proposed edit: Add a new section after Acknowledgments titled "Data Availability" with text such as: "Raw measurement data and analysis scripts used to generate the figures in this work are available from the corresponding author upon reasonable request. [Alternatively: are deposited at DOI:xxxx with CC-BY license.]"

Why this works: Satisfies PRX Quantum reproducibility requirements and addresses reviewer concerns about replicability.

4.8 majorana2 — Stage-6 CQE narrative

Model: `claude-haiku-4-5-20251001` · 1,089 output tokens · 28.3 s

Process Summary: QuantumNovelty Run Evaluation

4.8.1 Composite Verdict

The geometric mean composite score of **23 out of 100** places this run firmly in the **"Insufficient"** tier of collaboration quality. On the standard interpretation scale, scores below 30 indicate that fundamental workflow stages were either skipped or produced incomplete outputs, rendering the run unsuitable for any downstream scientific claims. This is not a marginal miss—a score of 23 reflects systemic gaps across nearly every dimension of the evaluation framework.

To be direct: this run did not produce a valid scientific artifact. The geometric mean methodology is intentionally punishing when dimensions cluster in the low range, because poor performance in any single area can invalidate an entire research effort. Here, no dimension exceeded 40, and multiple dimensions scored below 30, indicating that the collaboration failed to establish even baseline credibility checkpoints.

4.8.2 Strongest Dimension: Communication (Score: 40)

The **Communication** dimension achieved the highest score at 40, though this must be interpreted with significant caveats. Both constituent probes—"logical fallacies absent" and "reviewer panel verdict"—scored 40, but the evidence reveals why: neither validation step actually executed. The logical_fallacies skill was not run, and no review_panel.md file was generated.

This creates an evaluation artifact where the absence of detected problems is conflated with the absence of problems themselves. A score of 40 in Communication does not indicate strong scientific prose or clear argumentation—it indicates that the tools designed to surface communication failures were never invoked. The run cannot claim communication strength; it can only claim that communication quality remains untested.

If this is the strongest dimension, it speaks to the run's fundamental incompleteness rather than any particular competence. Communication typically improves as artifacts accumulate and undergo revision cycles. Without a paper draft (paper.tex exists: False), there was nothing substantive to evaluate for logical coherence or reviewer readiness.

4.8.3 Weakest Dimension: Novelty Rigour (Score: 8)

The **Novelty Rigour** dimension scored a critically low 8, driven by near-total failure in the baseline comparison infrastructure. The probe "augmented baseline catalog present" shows the baseline_catalog has **0 rows**—meaning no prior work was indexed against which to evaluate novelty claims. The "strict-domination comparator run" probe scored 5 because novelty_verdict.json was not found, indicating the comparator stage never executed or never completed.

This failure almost certainly originated in the **Literature Grounding stage** or its immediate successors. A QuantumNovelty run must first establish what constitutes known work before claiming any result is new. With an empty baseline catalog, no downstream novelty claim can survive scrutiny. Even if experimental results were generated (which other probes suggest they were not), they would be scientifically meaningless without situating them against existing methods.

The low Novelty Rigour score is particularly damaging because it undermines the entire purpose of the workflow. Quantum computing research is dense with incremental advances; claiming novelty without rigorous baseline comparison is not merely incomplete—it risks embarrassment upon peer review.

4.8.4 Three Highest-Leverage Improvements

1. Populate the Baseline Catalog Before Any Experimentation The single most impactful fix is ensuring the baseline_catalog contains relevant prior work before experiments begin. This is a gating dependency. The next run should include an explicit checkpoint that halts progression if baseline_catalog rows equal zero. Recommended implementation: integrate a mandatory literature sweep using established databases (arXiv, PubChem, prior QuantumNovelty archives) with human verification of at least 10-15 relevant baseline entries.

2. Generate Core Artifacts: paper.tex and audit_claims.py The Reproducibility dimension (score: 20) failed because foundational artifacts were missing. The paper draft serves as the canonical summary of claims, methods, and results; without it, there is nothing to audit. The audit_claims.py script ensures that every quantitative claim in the paper can be traced to source data. The next run should treat paper.tex generation as a mid-run checkpoint, not a final deliverable. Draft early, audit continuously.

3. Execute Domain-Specific Disclosures Explicitly The Domain Depth dimension (score: 30) failed across all three probes: no explicit active-space statement, no fermion-to-qubit mapping reference, and no simulator precision floor disclosure. These are not stylistic preferences—they are minimum credibility requirements for quantum chemistry claims. The next run should include a domain_disclosures.md template that forces explicit answers to these questions before any simulation executes. This prevents downstream reviewers from dismissing results due to unstated assumptions.

Bottom line: This run produced no defensible scientific output. The path forward requires treating artifact generation as mandatory infrastructure, not optional polish.