

## Peer-Review

Shen, Brandon, and Karena Ling. 2026. "Gradient Usability in Few-Qubit Quantum Neural Networks: A Signal-to-Noise Ratio Framework for Evaluating Mitigation Strategies." *Journal of High School Science* 10 (3): 358–93. <https://doi.org/10.64336/001c.166201>.

The manuscript proposes a hybrid ansatz framework based on the premise that constrained entanglement, local cost functions, and residual connections act through independent and complementary mechanisms. However, this premise is not substantiated. In variational quantum circuits, gradient behavior arises from coupled dependencies between circuit structure, measurement observables, and depth, implying that these interventions are not functionally independent when combined. The manuscript does not analyze interaction effects, cross-terms in gradient variance, or potential trade-offs in expressivity, sampling complexity, and hardware cost. Furthermore, in the few-qubit regime considered, true barren plateau scaling cannot be observed, limiting the ability to validate mitigation claims. As a result, the proposed framework remains speculative and does not meet the criteria for a review article that extracts new, non-obvious insights or proposes implementable solutions.

However, if you are open to 180 revision, you can frame your paper around 'usable gradients'. This means that you don't have the problem of whether you are observing plateau onset or true plateau (in your few qbit regime). Furthermore, you don't need to avoid variance collapse; just so long as the gradient variance can be distinguished from measurement noise. You can hence use the gradient SNR as a unifying metric. This reframes barren plateau mitigation as a problem of maintaining usable gradient signal. See the attached chatgpt file on how to design a falsifiable experiment along these lines.

You will have to also substantially rewrite the manuscript since you will now be deviating 180 from your current iteration.

---

## Response to Reviewer Comments

Manuscript: Ansatz Design for Mitigating Barren Plateaus in Few-Qubit QNNs: An Efficiency-Driven Approach

Authors: Brandon Shen, Karena Ling May 2026

The authors thank the Editor, Dr. Shireesh Apte, and both reviewers for their thorough and constructive engagement with the manuscript. The reviews identified substantive issues that had materially strengthened the revised work. Below, each reviewer comment is reproduced verbatim, followed by a description of how and where the manuscript had been revised in response.

## Reviewer 1

### Reviewer Comment 1.1:

*The manuscript proposes a hybrid ansatz framework based on the premise that constrained entanglement, local cost functions, and residual connections act through independent and complementary mechanisms. However, this premise is not substantiated. In variational quantum circuits, gradient behavior arises from coupled dependencies between circuit structure, measurement observables, and depth, implying that these interventions are not functionally independent when combined. The manuscript does not analyze interaction effects, cross-terms in gradient variance, or potential trade-offs in expressivity, sampling complexity, and hardware cost.*

### Response:

The reviewer's critique had been accepted in full. The original Section 5 had proposed a hybrid framework and implicitly treated the three components as independently beneficial — an

assumption the reviewer correctly identified as unsubstantiated given the coupled nature of gradient behavior in variational circuits.

In the revised manuscript, this assumption had been explicitly rejected. Section 5 had been entirely rewritten as a falsifiable experimental framework. Rather than asserting that the three components combine constructively, the revised section treats their interaction as an open empirical question. Three specific interaction pathways had been identified and analyzed:

- Constrained entanglement and local cost functions both act on gradient signal magnitude, meaning their combined effect on SNR may be sub-additive rather than multiplicative (Section 5.1.1, Interaction A).
- Residual connections may reintroduce correlations that entanglement constraints are designed to limit, creating a potential conflict between gradient preservation and entanglement control (Section 5.1.2, Interaction B).
- Residual connections and local cost functions operate in non-overlapping depth regimes, meaning their combined SNR benefit may be configuration-dependent rather than universal (Section 5.1.3, Interaction C).

A schematic gradient variance decomposition (Eq. 4) had been included to illustrate why the interaction between constrained entanglement and local cost functions involves a cross-term that cannot be decomposed as a product of individual effects. The decomposition had been stated to hold under the assumption that the circuit's output state satisfies area-law entanglement scaling — not a product state assumption, which would incorrectly imply zero entanglement — so that the approximation remains consistent with the paper's premise of constrained rather than eliminated entanglement. A caveat had been added that this decomposition is an approximation and that a rigorous derivation using the dynamical Lie algebra framework of Ragone et al. (2024) is identified as a direction for future theoretical work.

Trade-offs in expressivity, sampling complexity, and hardware cost had been addressed explicitly in the revised Table 1 (SNR effects column) and in Section 5.4 (Trade-offs and Practical Considerations).

**Location of changes:** Section 5 (entirely rewritten); Table 1 (SNR effect column added); Section 5.4 (new subsection).

## Reviewer Comment 1.2:

*Furthermore, in the few-qubit regime considered, true barren plateau scaling cannot be observed, limiting the ability to validate mitigation claims. As a result, the proposed framework remains speculative and does not meet the criteria for a review article that extracts new, non-obvious insights or proposes implementable solutions.*

### Response:

This concern had been addressed directly and had motivated the most significant structural change to the manuscript. The revised paper explicitly acknowledges that exponential gradient scaling as defined in Eq. 1 cannot be directly observed across a sufficient qubit range in 2–10 qubit systems. A paragraph in Section 1.2 now states this limitation explicitly and explains its consequences for evaluation.

The SNR framing had been adopted as the unifying metric throughout. Rather than asking whether gradient variance scales sub-exponentially with qubit count — a question that cannot be answered in the few-qubit regime — the revised manuscript asks whether gradient signals remain distinguishable from measurement noise under finite sampling. This question is directly answerable in 2–10 qubit systems and directly relevant to NISQ deployment.

Gradient SNR had been formally defined in Section 1.2 (Eq. 2) as the ratio of expected gradient magnitude to the standard deviation of the shot-noise estimator. The operability threshold of  $\text{SNR} \geq 1$  at  $N = 1000$  shots had been grounded in a first-principles derivation: for a bounded observable with eigenvalues in  $[-1, 1]$ , the standard deviation of a single-shot estimator is at most 1; under

parameter-shift estimation with  $N$  shots, the shot-noise standard deviation is therefore at most  $1/\sqrt{N}$ . At  $N = 1000$  this gives a noise floor of approximately 0.032, meaning  $\text{SNR} \geq 1$  requires  $\|E[\partial\theta C]\| \geq 0.032$ . Arrasmith et al. (2021) is cited only to support the subsidiary claim that gradient magnitudes above this floor correspond to meaningful optimizer progress — not as the sole basis for the threshold itself. The limitations of SNR as a proxy for trainability had also been stated explicitly: it is a necessary but not sufficient condition, and does not capture loss landscape geometry, optimizer dynamics, or parameter correlation.

The framework in Section 5 had been redesigned around SNR as the primary outcome measure, with a fully specified experimental design (Section 5.2), reference circuit configuration, and three falsifiable hypotheses (Section 5.3). Section 5.5 clarifies the scope of inference: results should be interpreted as characterisations of gradient concentration onset and signal degradation under practical constraints, and claims about asymptotic barren plateau mitigation should not be extrapolated from few-qubit results.

**Location of changes:** Section 1.2 (SNR definition; first-principles derivation of the  $\text{SNR} \geq 1$  operability threshold; explicit acknowledgment of few-qubit observability limitation); Section 5 (rewritten around SNR); Section 5.5 (new subsection on limitations of few-qubit evaluation).

Reviewer Comment 1.3:

*However, if you are open to a 180 revision, you can frame your paper around 'usable gradients'. This means that you don't have the problem of whether you are observing plateau onset or true plateau (in your few qubit regime). Furthermore, you don't need to avoid variance collapse; just so long as the gradient variance can be distinguished from measurement noise. You can hence use the gradient SNR as a unifying metric. This reframes barren plateau mitigation as a problem of maintaining usable gradient signal. See the attached chatgpt file on how to design a falsifiable experiment along these lines.*

**Response:**

This suggestion had been adopted in full. The manuscript had been substantially rewritten around the usable gradient framing. The abstract, introduction, and all analytical sections now foreground gradient SNR as the primary metric, and the experimental framework in Section 5 had been redesigned as a falsifiable experiment along the lines suggested.

Specifically, the abstract had been revised to state explicitly that SNR had been adopted as a unifying metric and that the question of whether mitigation strategies are effective had been recast as whether they preserve usable gradient signal, independent of whether true asymptotic barren plateau scaling can be directly observed in few-qubit regimes. Section 1.2 introduces SNR formally (Eq. 2), explains its motivating rationale, states both its utility and its limitations as a proxy, and grounds the operability threshold in a first-principles derivation. The phrase 'usable gradient signal' and its SNR operationalization now appear throughout the manuscript as the consistent framing device.

The experimental framework in Section 5.2 had also been made fully implementable by specifying a concrete target task: ground-state energy estimation for a 4-site open-boundary transverse-field Ising model with Hamiltonian  $H = -J\sum_i Z_i Z_{i+1} - h\sum_i X_i$ ,  $J = 1$ ,  $h = 0.5$ . This task had been selected because its local Hamiltonian structure is tractable on 4 qubits, its ground state is known analytically (enabling validation), and it is a standard benchmark in the variational quantum algorithm literature (Cerezo et al., 2021). The global cost function is defined as  $C_{\text{global}} = \langle H \rangle$  and the local cost function as  $C_{\text{local}} = \langle Z_0 Z_1 \rangle$ , the two-body term on the first qubit pair, consistent with the local observable construction of Cerezo et al. (2021).

The design recommendation in Section 2.3 had also been revised to include the caveat that the preference for problem-inspired ansätze should be validated empirically for specific tasks, rather than assumed from existing theoretical results. In the original manuscript, this recommendation was stated without qualification; the revised version specifies that the advantage had been formally

demonstrated only for translationally invariant QCNN circuits and hierarchical tree tensor network ansätze, and requires empirical validation for other problem-inspired designs.

A central analytical finding had also been added to Section 6: the conclusion now opens by identifying sub-additive SNR as the paper's primary non-obvious insight — that deploying constrained entanglement and local cost functions together does not yield multiplicative SNR improvement because both interventions act on the same underlying quantity, the distribution of quantum states the circuit can explore. This finding does not follow from examining either strategy in isolation and had not been previously identified in the literature.

**Location of changes:** Abstract (rewritten); Section 1.2 (SNR definition, framing, and threshold derivation); Sections 2.1, 2.2, 2.3 (SNR perspective integrated; design recommendation qualified); Table 1 (SNR effect column); Section 5.2 (target task and cost functions specified); Section 5 (entirely rewritten as falsifiable experimental framework); Section 6 (conclusion opens with sub-additive SNR as non-obvious finding).

Reviewer 2

**Reviewer Comment 2.1:**

*Please verify that all links to the references point to the correct source.*

**Response:**

All references in the revised manuscript had been verified. Each DOI or URL had been checked against the cited source. The following corrections had been made relative to the original submission:

- Reference [10] in the original version cited only the arXiv preprint of Kashif & Al-Kuwari. In the revised manuscript this had been updated to cite the published EPJ Quantum Technology version (Kashif & Al-Kuwari, 2024, doi: 10.1140/epjqt/s40507-023-00216-8) as the primary citation, with the preprint retained as a secondary reference where relevant. The revised manuscript cites the published version throughout.
- Reference [22] in the original version cited a 2024 arXiv preprint of the Larocca et al. review (arXiv:2405.00781). In the revised manuscript this had been updated to the published version: Larocca et al., Nature Reviews Physics, 7:174–189, 2025, doi: 10.1038/s42254-025-00813-9.
- Reference [13] in the original version cited an arXiv preprint of Liu et al. (arXiv:2112.10952). In the revised manuscript this had been updated to the published version: Liu et al., New Journal of Physics, 25:013039, 2023, doi: 10.1088/1367-2630/acb58e.
- Reference [14] in the original version cited an arXiv preprint of Kulshrestha & Safran (arXiv:2204.13751). In the revised manuscript this had been updated to the published conference proceedings version: Kulshrestha & Safran, 2022 IEEE International Conference on Quantum Computing and Engineering (QCE), pages 197–203, doi: 10.1109/QCE53715.2022.00039.
- Reference [20] in the original version cited an arXiv preprint (arXiv:2403.04844) for the barren-plateau-free HEA result. In the revised manuscript this had been updated to the published version: Park & Killoran, Quantum, 8:1239, 2024, doi: 10.22331/q-2024-02-01-1239.
- All remaining references had been confirmed to link to the correct source. No other corrections were required.

**Location of changes:** Reference list (multiple references updated from arXiv preprints to published versions; all entries verified).

**Reviewer Comment 2.2:**

*The authors must disclose and acknowledge any assistance received in the preparation of this manuscript, including but not limited to editorial, technical, analytical, or writing support. All such contributions must be clearly stated in the Acknowledgments section.*

**Response:**

The Acknowledgements section has been verified to explicitly disclose any assistance received, including the use of large language model assistants.

**Reviewer Comment 2.3:**

*Include enough recent references along with foundational ones. Ensure references directly support your claims. Avoid "padding." Use credible sources (peer-reviewed journals, reputable books, official reports)*

**Response:**

The reference list had been reviewed against these criteria. The following changes had been made:

- The published version of Larocca et al. (2025, Nature Reviews Physics) had replaced the preprint, providing an up-to-date comprehensive review of the barren plateau literature. Ragone et al. (2024, Nature Communications), which provides the unified dynamical Lie algebra framework for barren plateaus, had been added as Reference [6] in the revised manuscript and cited in Sections 1.2 and 5.1 where the theoretical underpinning of interaction effects is discussed.
- Park and Killoran (2024, Quantum) had been added to support the claim in Section 2.1 that an HEA construction had been demonstrated to achieve a constant lower bound on gradient magnitudes at any circuit depth.
- References [13], [14], and [10] in the original version had been updated from arXiv preprints to their published peer-reviewed versions (see response to Comment 2.1).
- The original manuscript cited a separate arXiv preprint of Kashif & Al-Kuwari (Reference [10] in the original) alongside the published EPJ version. In the revised manuscript, the published version is the primary citation throughout; the preprint reference had been retained only where the specific preprint content was cited.
- All 30 references in the revised manuscript are from peer-reviewed journals or peer-reviewed conference proceedings. Foundational works (McClellan et al. 2018; Cerezo et al. 2021; Holmes et al. 2022) are retained alongside recent publications from 2023–2025.
- **Location of changes:** Reference list; in-text citations in Sections 1.2, 1.3, 2.1, and 5.1

**Reviewer Comment 2.4:**

*All assumptions (including implicit assumptions) must be explicitly stated and clearly justified. The authors should explain why each assumption is reasonable and discuss its impact on the results and conclusions.*

**Response:**

Four assumptions that had been implicit in the original manuscript had been made explicit and justified in Section 1.2 of the revised manuscript. These are:

- **Scope assumption:** The analysis is scoped to 2–10 qubits on the grounds of tractability and relevance to current NISQ hardware. It is acknowledged that this range does not capture the full scale of deployed NISQ devices, and that conclusions may not generalize to larger systems.
- **Proxy assumption:** Gradient SNR is adopted as a proxy for trainability on the grounds that a gradient signal distinguishable from measurement noise is a necessary condition for effective optimization. It is explicitly noted that this is not a sufficient condition, and that other factors including loss landscape geometry, optimizer dynamics, and parameter correlation are not captured.
- **Strategy selection assumption:** The three candidate mitigation strategies examined in Section 5 had been selected because each targets one of the three primary barren plateau mechanisms identified in the literature, each had been empirically evaluated in isolation in prior work, and

together they span structural, functional, and architectural levels of circuit design. It is noted that this selection does not imply they constitute the optimal or complete set.

- Representational capacity assumption: Area-law entanglement constraints are assumed to preserve sufficient representational capacity for the tasks considered. The conditions under which this assumption is most and least likely to hold had been discussed explicitly, including the class of tasks for which it may fail (e.g., fully-connected combinatorial optimization problems requiring global correlations).

**Location of changes:** Section 1.2 (four explicit assumption paragraphs added); Section 5.5 (scope of inference clarified).

### **Reviewer Comment 2.5:**

*Avoid overreaching conclusions that extend beyond what is supported by the data and analysis.*

#### **Response:**

The revised manuscript had been reviewed systematically for overreaching claims. The following specific changes had been made:

- The design recommendation in Section 2.3 had been revised. The original version stated that problem-inspired ansätze "should be preferred" for "superior gradient scaling" without qualification. The revised version specifies that this advantage had been formally demonstrated only for translationally invariant QCNN circuits and hierarchical tree tensor network ansätze, and that empirical validation in the specific few-qubit task is necessary before adopting this preference for other problem-inspired designs.

- The conclusion section had been revised to remove the claim that structured ansatz design "consistently outperforms random initialization." This claim had been replaced with the more accurate statement that structured, informed circuit design had demonstrated more favorable gradient information preservation than random initialization across the architectures and settings reviewed.

- Section 5.5 had been added to explicitly state that the proposed experimental framework produces characterisations of trainability proxies, not definitive mitigation of asymptotic barren plateaus, and that results should not be extrapolated beyond the 2–10 qubit regime.

- A note had been added in Section 2.2 clarifying that QCNN barren plateau absence depends on the translationally invariant structure of the circuit and may not hold for problem-inspired designs lacking this symmetry.

- The limitation note in Section 5 had been retained and strengthened: the interaction between residual connections and constrained entanglement in the bounded Hilbert space of few-qubit systems is explicitly flagged as potentially counterproductive if skip connections inadvertently reintroduce entanglement that structural constraints were designed to limit.

**Location of changes:** Section 2.3 (design recommendation qualified); Section 6 (conclusion language revised); Section 5.5 (scope of inference); Section 2.2 (QCNN generalisability caveat); Section 5 (limitation note strengthened).

### **Reviewer Comment 2.6:**

*Verify that you have used past perfect tense and third person throughout the manuscript wherever applicable.*

#### **Response:**

The manuscript had been reviewed systematically for tense and person consistency. All references to prior literature had been revised to use the past perfect tense throughout (e.g., "had demonstrated," "had been identified," "had shown"). All first-person constructions had been replaced with third-person formulations. The following types of change had been made throughout:

- First-person phrases such as "we propose," "we survey," and "we identify" had been replaced with "a structured review is presented," "four approaches are surveyed," and "key gaps are identified."
- Simple past references to prior work (e.g., "Cerezo et al. showed") had been revised to past perfect (e.g., "Cerezo et al. had shown") to reflect that the cited findings preceded the writing of this manuscript.
- Present-tense generic statements about circuit behavior (e.g., "random circuits produce") had been retained where they describe mathematical properties rather than historical findings, consistent with standard practice in physics and engineering literature.

**Location of changes:** Throughout the manuscript, concentrated in Sections 1–4 and 6.

The authors believe the revised manuscript fully addresses all reviewer concerns. The substantive reframing requested by Reviewer 1 has been implemented throughout the paper, and all technical, stylistic, and disclosure requirements raised by Reviewer 2 have been satisfied.

---

The revisions substantially improve the manuscript by reframing trainability in the few-qubit regime around usable gradients and gradient signal-to-noise ratio (SNR). In my view, this reframing now represents the manuscript's most significant conceptual contribution. The argument that asymptotic barren plateau scaling cannot be directly established in 2–10 qubit systems, and that trainability should therefore be evaluated through gradient usability under finite sampling, is well motivated and addresses a genuine limitation of prior discussions.

However, it is less clear what additional contribution is provided by the proposed "sub-additive SNR" framework. The manuscript argues that mitigation strategies may exhibit non-additive interactions because they operate through overlapping mechanisms. While this is plausible, the manuscript explicitly characterizes these interaction effects as conjectures and hypotheses rather than established results. No theoretical derivation or empirical evidence is provided demonstrating that sub-additive behavior occurs in practice.

Consequently, I encourage the authors to reconsider the relative emphasis placed on this concept. At present, the SNR-based reframing appears to be the manuscript's primary intellectual contribution, whereas the interaction framework functions more naturally as a research agenda and experimental roadmap. The manuscript would benefit from clarifying whether "sub-additive SNR" is intended as a demonstrated insight, a motivating hypothesis, or a future research question. If the latter, the language in the conclusion should be adjusted accordingly to avoid giving the impression that the interaction effect has been established.

More broadly, the authors may wish to articulate more explicitly why the interaction framework constitutes a non-obvious contribution beyond standard ablation methodology. An informed reader might naturally infer that mitigation strategies should eventually be studied in combination. Therefore, the novelty of the interaction agenda would be strengthened by a clearer explanation of why existing intuition is insufficient, why overlapping mechanisms are expected to matter, or why the SNR framing reveals a previously overlooked aspect of trainability.

2. Present a stronger literature based chain of evidence positioning SNR as the relevant metric in the low qbit  $<10$  regime. Right now, the argument is : Few-qubit systems cannot establish asymptotic scaling....Therefore SNR is more useful.

Present this argument as a series of literature based steps:

- a) Show that optimization success depends on usable gradients: Cite papers showing that optimization fails when gradients become too small, regardless of whether a formal barren plateau has been established.
- b) Show that finite-shot estimation introduces a practical noise floor: Cite parameter-shift and NISQ literature demonstrating that gradients are estimated from finite samples and therefore have measurement uncertainty. Then argue that the optimizer does not see the true gradient; it sees a noisy estimate. This immediately makes SNR relevant.
- c) Show that few-qubit studies rarely probe asymptotic behavior because exponential scaling cannot be robustly inferred from three or four small system sizes. Therefore many "few-qubit

trainability” studies are not actually measuring barren plateau scaling.

d) Connect {a,b and c}

Then the synthesis becomes:

If optimization depends on usable gradients (Step 1), and gradients are estimated under finite sampling (Step 2), and few-qubit studies cannot directly establish asymptotic scaling (Step 3), then trainability in the few-qubit regime is more directly characterized by whether the gradient remains distinguishable from measurement noise than by asymptotic variance scaling itself.

You can claim a chain of evidence that shows that SNR may be a more operationally relevant metric for evaluating trainability in few-qubit systems where asymptotic scaling cannot be directly established. That is a much stronger and more defensible position. It turns SNR from an available replacement metric into a logically motivated consequence of the constraints of the few-qubit regime.

---

## Response to Reviewer Comments

### **Manuscript: Gradient Usability in Few-Qubit Quantum Neural Networks: An SNR-Based Framework for Evaluating Mitigation Strategies**

Authors: Brandon Shen and Karena Ling July 2026

The authors thank the reviewer for the careful reassessment of the manuscript and for identifying the finite-shot SNR reframing as its strongest conceptual contribution. The reviewer raised two principal concerns: the interaction framework was described more strongly than its evidentiary basis permitted, and the motivation for using SNR in the few-qubit regime required a clearer literature-grounded chain of reasoning. The manuscript has been revised throughout to address both concerns, distinguish established results from hypotheses, and remove technical ambiguities that could otherwise make the proposed experiment difficult to interpret.

#### **Reviewer 1**

##### **Reviewer Comment 1.1**

*However, it is less clear what additional contribution is provided by the proposed “sub-additive SNR” framework. The manuscript argues that mitigation strategies may exhibit non-additive interactions because they operate through overlapping mechanisms. While this is plausible, the manuscript explicitly characterizes these interaction effects as conjectures and hypotheses rather than established results. No theoretical derivation or empirical evidence is provided demonstrating that sub-additive behavior occurs in practice.*

#### **Response**

The reviewer is correct. The manuscript provides neither empirical evidence nor a theoretical proof that the complete estimator-SNR interaction is sub-additive. We have revised the manuscript to separate the following claims:

- For a local objective written as a weighted sum of fixed local observables, Equation 4 is an exact identity for the corresponding exact gradient. It shows how the entangling architecture changes reduced-state derivatives and how the local objective selects and weights the components that contribute to the measured gradient.
- Because the local terms can reinforce or cancel one another, Equation 4 does not determine the sign of an interaction. It provides a mechanistic reason to test the restricted-entanglement  $\times$  local-cost interaction at the exact-gradient-signal level. A negative coefficient, corresponding to diminishing marginal improvement on the prespecified scale, is treated only as one plausible outcome and not as a result established by the identity.
- Equation 4 does not describe finite-shot estimator variance. The complete estimator-SNR interaction can therefore be negative, negligible, or positive regardless of the exact-signal interaction. H1 has consequently been changed from a one-sided test for sub-additivity to a two-sided matched factorial test of the exact-gradient interaction coefficient. A negative coefficient may be interpreted as diminishing marginal improvement on the prespecified scale, but no sign is assumed. H2–H4 are also two-sided. The abstract, interaction discussion, hypotheses, limitations,

and conclusion now state consistently that all interaction outcomes remain unestablished until the proposed experiment is performed.

**Location of changes:** Abstract, Section 4.1, Section 5.1, Section 5.3, Section 5.5, and Section 6.

### **Reviewer Comment 1.2**

*Consequently, I encourage the authors to reconsider the relative emphasis placed on this concept. At present, the SNR-based reframing appears to be the manuscript's primary intellectual contribution, whereas the interaction framework functions more naturally as a research agenda and experimental roadmap. The manuscript would benefit from clarifying whether "sub-additive SNR" is intended as a demonstrated insight, a motivating hypothesis, or a future research question. If the latter, the language in the conclusion should be adjusted accordingly to avoid giving the impression that the interaction effect has been established.*

### **Response**

We agree with this hierarchy of contributions. The revised manuscript explicitly identifies the SNR reframing as the primary conceptual contribution and the interaction framework as a secondary, falsifiable research agenda enabled by that reframing. It also states that SNR is not introduced as a new statistical quantity; the contribution is the synthesis that places pointwise finite-shot estimator SNR at the center of few-qubit gradient-usability evaluation while retaining exact-gradient landscape variance as a separate theoretical diagnostic.

The abstract and conclusion now state that the review provides a literature-grounded reason to evaluate gradient resolvability through pointwise SNR, but does not demonstrate an interaction outcome, determine the sign of a complete estimator-SNR interaction, prove the absence of barren plateaus, or establish a universal optimizer threshold.

**Location of changes:** Abstract, Section 1.2.1, Section 5, and Section 6.

### **Reviewer Comment 1.3**

*More broadly, the authors may wish to articulate more explicitly why the interaction framework constitutes a non-obvious contribution beyond standard ablation methodology. An informed reader might naturally infer that mitigation strategies should eventually be studied in combination. Therefore, the novelty of the interaction agenda would be strengthened by a clearer explanation of why existing intuition is insufficient, why overlapping mechanisms are expected to matter, or why the SNR framing reveals a previously overlooked aspect of trainability.*

### **Response**

The revised manuscript no longer claims that the literature implicitly assumes additivity. The cited studies generally evaluated the selected interventions separately, which leaves the corresponding interaction terms uncharacterized but does not imply that prior authors asserted multiplicative gains. The contribution is now stated more narrowly. Its novelty is not the general suggestion that mitigation strategies should be studied in combination. Rather, it is the prespecified framework for separating exact-gradient-signal effects from finite-shot estimator effects while identifying all factorial interactions under matched resource constraints. Ordinary ablation does not by itself specify:

- which finite-shot outcome should represent practical gradient usability;
- how exact-gradient signal, estimator mean, finite-shot variance, estimator bias, and sign agreement should be separated;
- which interaction has a mechanistic basis for prespecifying a plausible interpretation, while all confirmatory tests remain two-sided;
- which eight configurations are required to identify all pairwise and three-way effects independently;
- how total measurement budget, initialization, depth, task fidelity, and resource cost should be controlled; or
- how interaction estimates should be analyzed while propagating uncertainty from repeated finite-shot gradient batches.

The SNR framework supplies this organization. The proposed experiment uses a complete  $2^3$  factorial design, matched initializations and quantum-gate parameters, equal total measurement budgets, separate exact-signal and estimator-SNR models, nested matched bootstrapping, and family-wise error control. The cost-function coefficient is explicitly scoped to the comparison between normalized TFIM energy and global target-state infidelity rather than attributed to locality alone.

**Location of changes:** Section 4.1, Sections 5.1–5.4, Section 5.5, and Section 6.

#### **Reviewer Comment 1.4**

*Present a stronger literature based chain of evidence positioning SNR as the relevant metric in the low qbit  $<10$  regime. Right now, the argument is: Few-qubit systems cannot establish asymptotic scaling; therefore SNR is more useful. Present this argument as a series of literature based steps: (a) show that optimization success depends on usable gradients; (b) show that finite-shot estimation introduces a practical noise floor and that the optimizer observes a noisy estimate rather than the true gradient; (c) show that few-qubit studies rarely probe asymptotic behavior because exponential scaling cannot be robustly inferred from only a few small system sizes; and (d) connect these steps to motivate SNR as an operationally relevant metric in the few-qubit regime.*

#### **Response**

This recommendation has been implemented in the dedicated subsection “Why Gradient SNR Is an Operationally Relevant Metric in the Few-Qubit Regime,” organized as the requested three-step chain and synthesis.

1. Step 1 explains that optimization requires sufficiently precise and directionally useful gradient information, not merely the mathematical existence of a nonzero derivative. Arrasmith et al., the VQA reviews by Cerezo et al. and Larocca et al., and the adaptive-shot work of Tamiya and Yamasaki support this step. The manuscript explicitly avoids claiming a universal gradient-magnitude threshold.
2. Step 2 explains that parameter-shift gradients are estimated from finite samples and therefore have configuration-dependent sampling uncertainty. The optimizer receives an estimator rather than the exact derivative. The standard deviation scales as  $O(N^{-1/2})$  under a fixed estimator and allocation rule, but the proportionality constant depends on the state, observable decomposition, shift coefficients, grouping, and budget convention. The text therefore uses “uncertainty scale” rather than an irreducible universal noise floor. It also specifies that the fraction of gradient components with estimator  $\text{SNR} \geq 1$  is only a descriptive resolvability benchmark, not a necessary or sufficient threshold for optimization success.
3. Step 3 distinguishes finite-size characterization from asymptotic classification. Measurements over 2–10 qubits can characterize gradient concentration, estimator uncertainty, and optimization behavior, but generally cannot independently determine an exponential or polynomial asymptotic class without uncertainty estimates, competing-model comparison, and theoretical support. Clauset et al. and Harada are cited for empirical functional-form and finite-size-scaling methodology, alongside the foundational barren-plateau literature.
4. The synthesis states that, because optimization depends on usable estimates, those estimates are obtained under finite sampling, and short few-qubit sweeps generally do not determine an asymptotic class by themselves, pointwise estimator SNR addresses the immediate experimental question of whether the optimizer can resolve the update signal. Exact-gradient landscape variance remains a complementary barren-plateau diagnostic and is not replaced.

**Location of changes:** Section 1.2.1, Section 6, and References.

#### **Additional Technical and Presentation Revisions**

Several additional changes were made to make the proposed framework internally identifiable, reproducible, and interpretable:

- The residual shortcut used in the primary matched-initialization analysis is now active. The residual condition uses a fixed prespecified gain  $\gamma = 1$ , while the baseline uses  $\gamma = 0$ . A trainable gain initialized at zero is retained only as an optional longitudinal exploratory variant because a

zero gain would make the residual and baseline architectures functionally identical at the primary analysis point and would make H3 and H4 incapable of detecting residual effects.

- The former mixed-state-fidelity local objective has been replaced by normalized TFIM energy, expressed through fixed one- and two-qubit terms. The global objective is written as target-state infidelity using the target projector. Both objectives share the same nondegenerate target, but the manuscript now states explicitly that their contrast also differs in spectral weighting, term aggregation, and measurement decomposition.
- Equation 4 has been generalized from a single local observable to a weighted sum of local observables, matching the proposed TFIM objective. The text explicitly notes that terms can reinforce or cancel and therefore do not imply a universal interaction sign.
- The primary exact-signal H1 test is now two-sided. A negative coefficient may be interpreted as diminishing marginal improvement on the prespecified scale, but the confirmatory test does not assume that sign.
- The finite-shot protocol distinguishes the exact total derivative, estimator mean, sampling variance, estimator bias, and sign agreement. The primary analysis is restricted to quantum-gate parameters matched across all configurations. The fraction of gradient components with estimator  $\text{SNR} \geq 1$  is reported only as a descriptive resolvability benchmark and is not treated as a necessary or sufficient threshold for optimization success.
- Replicate count and matched-initialization count are selected through independent pilot-based precision rules. Pilot data used to select those counts are excluded from confirmatory analysis.
- Equal measurement budgets include every forward-feature estimate, shifted circuit evaluation, observable group, and node-level Jacobian needed by the complete hybrid computational graph.
- The manuscript now uses simple past or present perfect consistently in the literature review, shortens the abstract, removes unsupported categorical language, and clarifies all few-qubit and task-specific limitations.
- The submission file has been reformatted as a one-column DOCX in Times New Roman 12 point, 1.15 line spacing, with the required margins, blank lines between paragraphs, numbered equations, a formatted table, numeric references, and live DOI links.

**Location of changes:** Section 5.2.1, pp. 20–21; Sections 5.2.2–5.2.3, pp. 21–25; Section 5.3, pp. 25–

27; Section 5.5, p. 28; and throughout the manuscript.

These revisions clarify the hierarchy of contributions, strengthen the literature-grounded SNR rationale, separate established identities from interaction hypotheses, make all confirmatory interaction tests identifiable and appropriately two-sided, and provide a complete, prespecified experimental roadmap without claiming results that have not been demonstrated.

---

Thank you for addressing my concerns. Accepted.