nature portfolio

Peer Review File

Observation of constructive interference at the edge of quantum ergodicity

Corresponding Author: Dr Hartmut Neven

Any redactions in this file are there to maintain patient confidentiality, the confidentiality of unpublished data, or to remove third-party material.

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Referee #1

(Remarks to the Author)
In this work Google quantum AI reports on a series of measurements of out-of-time-order correlators (OTOCs) in random quantum circuits. These circuits are constructed to scramble quantum information slower than maximal, such that the OTOC can be estimated efficiently on the quantum device. It is argued that at sufficiently long times these results can not be obtained via known classical algorithms, of which several are developed by the authors themselves and extensively discussed in the supplementary information. As such, the authors claim to fulfill the longstanding goal of achieving quantum advantage in computing local expectation values. Finally, it's argued that these results are directly relevant for modeling of ESR and NMR spectroscopy of strongly correlated molecules; making the quantum advantage practical as claimed in the title.

There is of course no doubt that these results are based on impressive recent developments in Google's quantum technology. However, that is not the point of the paper so I feel obliged to judge it based on the novelty of its claims, not on previously reported developments (re Willow). At a high level, I simply find the paper tedious to read. For me, it reads like a gallimaufry of technicalities, e.g. the number of figures with different panels and captions is overwhelming. It seems a symptom of the lack of a clear crisp result that supports the claim of practical quantum advantage for local expectation values. That being said, I do want to stress that the authors make an honest and thorough attempt to develop the best possible classical method.

Some more specific comments are:

While several classical numerical methods are explored, I am unaware of any theoretical result that shows exponential separation between (noisy) quantum and classical computing on the problems discussed. In fact, following arXiv:2407.12768 it seems plausible that an extension to OTOC's would formally be polynomial albeit very impractical. This does beg the question whether some more refined approximate methods wouldn't outperform those of the authors. Possibly some adaptation of DOI: 10.1126/sciadv.adk4321.
To be able to actually estimate the OTOC efficiently on the quantum device, it can not be too small since it's estimated from the average outcome of a single qubit measurement. Slowering scrambling circuits are therefore proposed, but it's not clear to me how this affects the hardness. Naively, it would mean that Pauli strings also spread less quickly in the Pauli space and if this effect is parametrically different it should also make classical simulations more efficient. For example, it's not clear to me that this doesn't make it easier to predict the circuit-to-circuit fluctuations.
Error mitigation is used to correct for some effects of the noise. That this creates some uncertainty in the results can be seen from Fig 2b. The OTOC can not exceed 1 if it was estimated without any post processing. It's clear at 15-20 cycles qubits that are still outside the lightcone start to have some fluctuating OTOC which exceeds 1. I consider this as an estimator of the error bars on the mitigation. It's of course hard to estimate from the plot, but it's visible by the naked eye. Turning to panel 2e, it looks like these O(10-3) ought to be more affected by this mitigation procedure. The fact that they are not, suggests that these measurements are more robust, again begging the question whether they can not be computed more easily classically.
In a previous work Google already reported on OTOC extracted from an earlier device: DOI: 10.1126/science.abg5029. I find

that most of the discussion and the philosophy behind these measurements is very similar. So similar, that I am not sure what the added value of this discussion is.
I find the discussion of real-world applications, in particular the matching of quantum simulations to experimental data, a bizarre twist. It has barely anything to do with the results presented in the paper and there are 7 pages of supplementary material devoted to it. I look forward to seeing this implemented on Google's device, but since this was not reported in the paper I don't see the rationale.

In conclusion, the authors report on OTOC results on their new quantum processor. The results themselves are impressive, but not novel.

Referee #2

(Remarks to the Author)
None

Referee #3

(Remarks to the Author)
This paper introduces several new experimental and theoretical insights at the absolute forefront of quantum computing science. The authors report on the use of a 103-qubit superconducting quantum processor to measure out-of-time-order correlators (denoted OTOC and its higher-order generalization OTOC2). This new generation processor overall has improved performance relative to previous work from the Google Quantum AI team and is reaching over-more-impressive capabilities. In addition to measuring OTOC/OTOC2 in limit(s) that are challenging to simulate classically, they develop a new Monte Carlo-based algorithm to classically simulate the values of the OTOC. However, for OTOC2, which includes higher-order (quantum) interference effects, the authors estimate that reproducing the measured OTOC2 values (for their largest systems) is beyond state-of-the-art classical supercomputers even with modern developments of classical algorithms. Finally, the authors develop a protocol that probes whether quantum correlations play a role in the measured values of OTOC/OTOC2 and find that within their protocol there is indeed large contributions from quantum correlations.

Overall, this is (yet another) extremely impressive experiment and theory development from Google Quantum AI (and collaborators). The processor uses the same fundamental ideas as previous processor(s), but (as briefly mentioned in the supplement) re-designing the ground plane near the qubits has led to improved T1 times (now ~106μs median) which is the main driver of their improved hardware numbers. To the best of my knowledge this is – viewed holistically – one of, if not the, most performant quantum processor in the world. The continued insistence on pushing the frontiers of quantum information science with this (family of) processors by the Google Quantum AI team is very impressive and I deem the overall quality, impact, and importance of this particular work to be very much suitable for publication in Nature. It will set a new bar in state-of-the-art quantum computing - at the scale of ~100 highly performant qubits - pushing ever-closer to solving computational problems of practical importance, maybe even before full-scale quantum error-correction.

The primary features of this manuscript that warrants publication in Nature in my opinion is the following (in no particular order):
Developing a new method for 'demonstrating' that quantum interference effects play an important role in a dataset from a quantum processor at a scale beyond traditional tomographic or other means of recreation/QCVV certification is a crucial technique.
Developing a framework for relating OTOC-measurements to a real-world practical problem within NMR (although this development is no tthe main focus of the main manuscript, there is a lot of details on this connection in the supplement).
Exercising a ~100 qubit quantum processor and outputting quantities (values of OTOC / OTOC2) that seem to be significantly beyond what state-of-the-art classical supercomputing can currently achieve, given our best-known classical algorithms.

Overall I view the combination of hardware developments, conceptual and theoretical insights to be of the highest-impact in our field and thus consistent with a journal such as Nature.

With that said, I have some comments (below). They broadly fall into two categories: (1) technical questions that I think the authors should address and consider including their responses in the manuscript/supplemental to alleviate these concerns should other readers and (2) minor comments that I would propose to improve readability/clarity of exposition. My points in (1) are not fundamental technical concerns, but rather represent what I think could also be reasonable considerations other (careful) readers might also consider.

The readout on this processor has significantly improved (along with everything else, mostly due to T1, as the authors state). But what is the readout time? I couldn't find it anywhere. Consider including this, along with median, 1QB and 2QB gate times in the supplement.

Fig 2b: Here (and for fig. 2c) the authors write in caption "with different qubits as Q_b". I would encourage the authors to visualize this in the figure diagrams in some way. It is hard to build an intuition for the data with the vague description "different qubits as Q_b".

Fig 2c: For certain values (e.g. distance ~24 and cycle ~23) the avg. OTOC is larger than 1. What is the interpretation of this?

Fig 2d: The double-x axis for OTOC(s) and the TOC is confusing. Why are the TOC Cycle axis (top) chosen to go to 19? Why not to 10 where the last datapoint is, like OTOC axis? Perhaps this is to make the gray and purple decays 'similar', but is it perhaps misleading? Not a crucial point, but something to consider.

Fig 2d: I would propose to draw a 'TOC' circuit diagram somewhere (just in supplement would be fine)

Fig 2d: I think an inset showing the experimental setup (indicating Qm and Qb would be really helpful, like in Fig. 2e and f)

I couldn't find a description of how many qubits are involved in the data in Fig. 2d)

Fig. 2f: Regarding this data I think the authors should consider including a comment/discussion to explain why, even when averaged over paulis, there are still certain datapoints that have significant C^2 (x 10^3). My understanding of this data is that there are three different families of results:
Type 1: Circuits where outcome is dramatically reduced by averaging [e.g. circuit instance ~29, with value approx. -2.5 (/10^3) and circuit instance ~27 with -5.5 (/10^3)].
Type 2: Circuits where outcome is changed partly when averaging [e.g. circuit 46 and 48 where the averaged value is reduced, but has not gone to zero].
Type 3: Circuits where C^2 is enhanced when averaging (e.g. circuit ~7, where yellow point and black sim are at about 5 (/10^3).
In all cases, the impressive MC simulation clearly captures the averaged value (as it should)

- Based on the discussion in the main matter of the paper I find there to not really be a discussion of neither Type 2 nor Type 3 circuits. Perhaps the authors could clarify this? This discussion (again, in the supplement is totally fine) would help the reader navigate these types of plots.

- Is the intuition here that if one was to average enough then all circuits would be of Type 1? Or is there something intrinsic going on here?

I found the discussion surrounding Fig 3c in the main text oddly hard to follow. In particular, is the data in Fig.3c bottom (purple) just the difference between Fig. 3c top (red and blue curves)? If so, then I think standardizing the visual presentation of these datasets and replacing the legend in Fig. 3c bottom with some more explicit would be helpful. A sentence in the main text explaining why Paulis injected at middle of U (versus ~1/3 and ~2/3 through U) does not kill interference would also be helpful. Currently there is just a sentence in the paper saying it is so, but not why.

Fig 3c top (minor): Would it make sense to plot this data on a log scale? Just to help the reader.

Fig 3c bottom (minor): I believe the y-axis is mislabeled. Should have been ΔC^((2)) (Delta is missing)

On p.4 top left the authors estimate '3 days on Frontier […]'. I think the authors should add somewhere around here the wallclock time to sample the OTOC2 on the quantum processor.

On p.5: I found the introduction of the Loschmidt echo to be quite ad hoc. I would strongly urge authors to better motivate this choice. Even when looking in the supplement it still begs the question (at least to me): why are we introducing this fidelity metric? It was also used in a related context in a paper also on superconducting qubits and measuring OTOCs (Braumueller et al, Nat Phys 2021, which could be added as a citation). There is of course nothing wrong with the loschmidt echo fidelity, but readers might wonder why it is suddenly introduced here.

Fig 4e (minor): I found the legend hard to read. Perhaps replace (1), (2) and (3) with something like (105q), (72qv1), 72q(v2) or something like that. Obviously a very minor point.

In VI: Methods – Choice of initial states. Please refer directly to the point in the SI where the 66 qubit circuits which mimic OTOCs are (for other points that refer to SI you refer directly to the section(s) in the supplement that's relevant)

Supplementary Fig. S35 (minuscule): Typo in the title.

(Remarks on code availability)
For obvious reasons I cannot actually execute the code (I do not have a 65 qubit backend :-), but the code seems complete, albeit rather simple. There are many other circuits in the paper that seem to not be included with the code. The dataset in the 'practical quantum advantage limit' (Delta C4) is the only circuit in the code. Including code for smaller system (e.g. the 31 qubits involved in Fig. 3c,d) could be included such that a representative simulation could be run locally.

Referee #4

(Remarks to the Author)
The current experiment is impressive both in scale and accuracy. Measuring OTOCs is an experimental challenge. Most notably, the present experiment not only captures the mean, but also fluctuations and higher moments of the OTOC. Thus,

the present work appears to me as an impressive experimental advance.

Despite of this experimental progress, I see concerns relating to the presentation of the paper in its current form:

1. In order to show any practical quantum advantage, as argued in the manuscript, it is key to go beyond mere academic questions such as sampling from a random circuit. In the present work, OTOCs are measured, which contain more physics, such as about scrambling and thermalization dynamics. However, it is not so clear to me what one can actually learn in terms of physics from OTOCs and its higher-order variant OTOC^2. Although there has been significant theoretical research on OTOCs in quantum matter within the last decade, it has remained open to me for what kind of general physical question it is the OTOCs which could give us new insights (despite of the rather theoretical question of operator spreading). While I understand that in the outlook Hamiltonian learning in the context of NMR is mentioned (and further discussed in the supplementary), it is not clear to me that OTOCs are the key element and whether for instance other measurements could also allow for doing Hamiltonian learning. The manuscript would profit significantly from discussing the physical importance of OTOCs and the higher-order variant OTOC^2.

2. What I find particularly challenging to clearly assess positively is the claim about quantum advantage. It is very much appreciated that the present work puts a lot of effort in comparing the experimental results to outcomes of potential classical algorithms. When it comes to the numerically exact simulation of the circuits, the authors, however, take as the standard reference tensor network methods, and for concrete comparison one particular variant. It is not at all clear to me that this naturally exhausts potential candidates in theory and justifies the claim of (practical) quantum advantage.

While a lot of research in theory targets to solve 2D quantum dynamics with tensor networks, it is, however, not at all clear according to the current state of the art, whether tensor networks are the ideal reference candidate (as it is taken in the present manuscript). In fact, it is generally accepted that tensor networks will necessarily fail to describe thermalization dynamics in 2D. In this context a very promising alternative method, which for dynamics in 2D has already demonstrated to be often superior compared to tensor networks, is variational Monte Carlo, or in more general terms neural quantum states. Because of the recent advances of these Monte-Carlo inspired algorithms, I would suggest to revise the manuscript to properly account for this class of methods, when claiming quantum advantage.

3. Large parts of the manuscript aim at demonstrating the importance of quantum mechanical interference in the dynamics of the OTOCs. This is important because many aspects of thermalization dynamics can be actually described purely classically in terms of hydrodynamic equations of motion (this could be also mentioned in the main text as a limitation of claiming quantum advantage). I have to admit that the presentation in the manuscript about the precise setup and implications of the random injections of Pauli operators is not very clear to me. When comparing the data of the original OTOC and OTOC^2 to the ones with random Pauli injections, one can certainly see some quantitative changes in the results. However, it also appeared to me that there is no qualitative change. In order to actually settle whether quantum interference is of key importance, it would be necessary to study this systematically as a function of system size. Overall, it is not clear to me of how fundamental importance quantum interference actually is here, which, however, might also be rooted in not properly appreciating the random injections due to an unclear presentation.

4. In the outlook the manuscript speculates that the hardware development "will rapidly increase the margin of quantum advantage, outrunning any gains that may result from improving classical simulation algorithms". This speculative comment might be correct when taking as a reference tensor networks, but it might not be for alternative numerical methods, which don't suffer from entanglement growth or matrix contraction complexity. I would suggest to adapt this speculation accordingly.

Besides these more general concerns, there are a few further aspects which might be worthwhile to address in a revised version:

1. On page 2 it is speculated that due to the change of the spectrum of the correlation operator a phase transition is expected. Is this meant to be a phase transition in time or as a function of some other control paramter? Could it also be a (sharp) crossover instead?

2. On page 2 it is said that a particular classical simulation would require 10^25 years. To me it is not clear whether this number for this one chosen reference candidate can be seen as representative.

Overall, I consider the current experiment to represent an impressive experimental advance. Due to the abovementioned concerns, I currently cannot recommend publication of the manuscript in its present form.

(Remarks on code availability)
I have checked the code, that was provided to me on the portal. I have downloaded, read the instructions, and executed the program. Everything worked out. It is worthwhile to mention that the provided data only refers to Fig. 4f and thus only provides a potential check for this particular figure.

Version 1:

Reviewer comments:

Referee #1

(Remarks to the Author)
The paper presents an experiment on a 103-qubit superconducting quantum processor. The experiment consists of measuring a generalization of the out-of-time-order correlator to a multi-round setup. The results focus on the second-order version in which the echo sequence is repeated twice.

The paper has been substantially revised as compared to the previous version, and many of the concerns have been resolved. The results are more focussed and the presentation is more clear. The main claim of the authors is that "interference between Pauli strings that form large loops in configuration space" makes the classical simulation intractable. That such loops contribute is evidenced by checking the sensitivity of the results to a Pauli-insertion scheme which randomizes the signs of these loops, thus suppressing their contributions.

In the introduction the authors present very well the 3 criteria for a worthwhile demonstration of quantum advantage. I summarize, (i) identify observable with a signal-to-noise ratio that is large enough such that it enables efficient extraction on the noisy quantum device, (ii) the same observable must be sufficiently complex such that it eludes classical simulation, (iii) there needs to be some practical use for the observable.

In my opinion the paper falls flat of (iii), but that's a high bar so let's leave that out for now. My main concern is that there is a fundamental tension between (i) and (ii) that, in spite of the thorough analysis in the supplement of the paper, I do not see resolved. The fact that the otoc only decays algebraically, allows to satisfy (i) but possibly opens the door to classical simulation as the amount of information one needs to capture about the state of the environment is now fundamentally limited. Of course it might well be that no classical algorithm can efficiently find this optimal compression, but that remains to be seen. This leaves the door open for future algorithms to simulate the experiment.

Referee #3

(Remarks to the Author)
In this revised version of the paper from the Google Quantum AI team and their collaborators, they have made many deep changes to the paper. While the fundamentals of the original data is still present, several new important (and impressive and clarifying) datasets/analysis are also presented, overall dramatically improving the quality of the manuscript.

While the fundamental dataset related to 'experimental quantum advantage' (essentially, Fig.4) is not changed, the exposition and the arguments leading up to this dataset are now vastly clearer and the impact of the result is significantly stronger. The authors have most certainly taken into consideration both mine (and other referees) comments in these wide-ranging revisions. Overall, this overhaul of the paper dramatically increases its impact in my view: instead of working hard to phrase their result in terms of 'practical' quantum advantage, it is now a study of 'experimental' quantum advantage and on balance the presentation is now much clearer, and the impressiveness stands out better. In particular the argument around interference effects in OTOC(n) values is much more pedagogical (Fig3 a,b,c) and I think will help other researchers better appreciate the physics being used (and perhaps spur others to utilize this technique more, now that it's been clearly laid out). The impact of what is achieved is thus more apparent and easier to understand.

I applaud the authors for now carefully spelling out what they deem a 'current generation quantum advantage claim' must satisfy (I agree with their arguments here – others could reasonably disagree, but it's a quantitative attempt at defining something elusive). In this revised version of the manuscript, they then put their OTOC(n) idea and hardware to the test on this 'benchmark'. While details of when quantum advantage has been definitively achieved are a gray zone (as the authors are well aware and mention in the paper), this paper sets a benchmark for how the community could/should think about quantum advantage in current/near term processors. With the publication of this manuscript in a foremost journal such as Nature I think this approach to claiming 'quantum advantage' will (and should!) become an industry standard – a throwing down of the gauntlet for other teams in the 'current generation quantum advantage' game. The added explicit "definition" of a Goldilocks zone (Fig. 1a) I'm sure we'll be used as a helpful guide for other teams in this community as well.

So, I'm very happy to recommend this manuscript for publication in Nature in its current form. It is an impressive revision which better presents the impressive experiment and offers important insights on how/when the community can claim quantum advantage (beyond random circuit sampling style arguments).

(Remarks on code availability)
The code now contains many more examples and details as compared with previous submission.

Referee #4

(Remarks to the Author)
The authors have taken signficant efforts to improve the manuscript, which is very much appreciated. The presentation now is much clearer and easier to follow.

I'm also still impressed by the quality of the data and the device, enabling the study of complex objects such as what is called in the manuscript OTOC^2.

Nevertheless, I cannot recommend publication of the manuscript in Nature. As discussed in more detail below, I'm not convinced by the claim on experimental quantum advantage and by the practical relevance of the considered observables.

Let me in the following comment on the individal responses of the authors.

1. As the authors have adapted the definition of what they claim as quantum advantage, the critique formulated in my first report cannot be directly applied anymore. While the new definition appears reasonable, it certainly matches more closely the results presented in the manuscript. However, criterion iii) on revealing pratically relevant information is only considered theoretically and not demonstrated on the actual quantum device, the reason for which remains unclear. Taking the definition for experimental quantum advantage formulated in the manuscript literally, I don't see their experiment fulfilling the claim.

2. The authors have put forward substantial efforts to further strengthen their claim that the quantum dynamics studied on the quantum processor cannot be simulated efficiently at least for large system sizes. I understand that the key argument is that the classical algorithm would have to learn the intricate sign structure in the quantum states. It is then argued along various axes why the neural quantum states method cannot achieve this.

First let me point out that neural quantum states are now very successful in learning complex sign structures in frustrated quantum magnets and interacting fermion matter:

https://arxiv.org/abs/2311.16889
https://doi.org/10.1038/s41567-024-02566-1
https://arxiv.org/abs/2405.04472

To support their claim the authors included also some analysis in the supplementary material, which seems to indicate that the numerical effort of the neural quantum states method would be exponential in the number of qubits. However, I have quite significant doubts whether the presented results are actually meaningful. First, the performed supervised learning procedure uses adam as the optimizer, which can be stuck in local minima and which it is typically not utilized. Second, the artificial neural network output is split to yield separately magnitude and phase. It is by now well known that such a choice is not suitable to detect nontrivial sign structures. Instead, transformer or CNNs for the full amplitude are typically used. Therefore, it is clear that the author's analysis is not suitable to exclude the neural quantum states method.

Further, the references mentioned in the author's report on previous works have left out some of the most important contributions of nonequilibrium dynamics solved with neural quantum states:

https://doi.org/10.1103/PhysRevLett.125.100503
https://doi.org/10.1126/sciadv.abl6850
https://arxiv.org/abs/2412.05332
https://arxiv.org/abs/2412.11778

for up to 400 qubits. Let me note that these examples contain far-from-equilibrium dynamics far beyond what is manageable using tensor networks due to significant entanglement generation.

Let me emphasize that I consider the thorough analysis of an alternative method such as neural quantum states only important because of the strong claim of practical quantum advantage, which, as I think, makes it necessary to rule out any classical competitor.

3. In the revised manuscript the authors have significantly improved the presentation on the random injections, which is now much clearer. I very much appreciate the underlying idea to identify quantum interference effects in the dynamics. The new data in Fig. 3c now displays more convincingly that the random injections have a more profound impact of C^4 as compared to C^2. In order to make the presentation in this context finally compelling, it would be important to more quantitatively assess the analysis of the C^4. In the inset of Fig. 3c a fit is presented to extract what is called the Pearson correlation. From the data for C^4 this fit doesn't appear too meaningful due to the wide scattering of points. At least I'd suggest to include the error bars in Fig. 3c which emerge from this fitting. I could imagine that with error bars the difference between C^4 and C^2 might not be that significant anymore.

4. This point has been largely addressed. It might be just fair to also mention that quantum theory could also improve so that it is not clear whether the 'quantum advantage' window in Fig. 1 would always increase.

5. This aspect has been fully clarified.

6. I understand that the estimation of classical resources are based on reasonable assumptions. However, what I find difficult to assess in a quantitative fashion is to extrapolate these estimates in order to gauge what might be possible in the near future. For instance, if GPUs could be utilized more efficiently, many orders of magnitude can be easily gained in any of such estimates.

Version 2:

Reviewer comments:

Referee #4

(Remarks to the Author)
The manuscript has seen substantial modifications, which is very much appreciated. The restructuring of the presentation away from the claim of quantum advantage towards more addressing physical questions has led to a strong improvement. Overall, the revised version is now clear and accessible.

The experimental results I still consider truly impressive, being capable of experimentally accessing such subtle quantum interference effects in a large many-body system.

Thus, I can now strongly recommend the main manuscript for publication in Nature.

Concerning the authors' response and some parts of the supplement, however, I respectfully disagree.

The authors still seem to not properly present the existing literature and knowledge gained in recent years on neural quantum states, which is important insofar that any claim about advantages over classical algorithms is directly affected by this.

First of all, it is crucial to accept that quantum entanglement, which is critical for most tensor network approaches, is not a limiting factor for neural quantum states, see for instance:

https://doi.org/10.1103/PhysRevX.7.021021
https://doi.org/10.1103/PhysRevLett.122.065301
https://doi.org/10.1103/physrevb.106.205136

Thus, any statement setting a potential failure of neural quantum states into the context of entanglement, as done in the response and the supplement, doesn't appear viable (while I would like to apologize for providing an incorrect reference to time evolution in my previous report).

This becomes slightly more severe in view of the presentation of achieved advances in time evolution for neural quantum states in the supplement, where it is suggested that neural quantum states might be standing behind tensor networks. Actually, the considered works demonstrate clearly (also partly in direct comparison) that neural quantum states can go significantly beyond.

As a side remark, it remains important to point out that utilizing stochastic gradient descent based solvers for neural quantum states doesn't represent state of the art, but rather higher-order solvers such as stochastic reconfiguration are utilized because only then the approach can become numerically exact with well controlled error bars.

I would therefore strongly suggest to adapt these parts of the supplement accordingly.

Response to Referee #1

Comment: At a high level, I simply find the paper tedious to read. For me, it reads like a gallimaufry of technicalities, e.g. the number of figures with different panels and captions is overwhelming. It seems a symptom of the lack of a clear crisp result that supports the claim of practical quantum advantage for local expectation values. That being said, I do want to stress that the authors make an honest and thorough attempt to develop the best possible classical method.

Response: We agree with the referee that our initial manuscript was too laden with technical details and the main findings were obscured. The manuscript is now re-written almost completely and focuses primarily on OTOC$^{(2)}$. The figures and technical discussion have been simplified and sharpened to highlight three key findings related to the titular quantum advantage, namely: 1) The noiseless circuit-to-circuit variance of OTOC$^{(2)}$ decays polynomially rather than exponentially as typical in ergodic dynamics. 2) We observe experimentally and theoretically that OTOC$^{(2)}$ is dominated by many-body interference effects that cannot be classically simulated. 3) OTOC$^{(2)}$ is sensitive to microscopic details of quantum dynamics which makes it a practically useful tool toward Hamiltonian learning.

Many of the technical figure panels related to classical simulation complexity or previously studied first-order OTOC$^{(1)}$ have been delegated to the supplement. In their place, we have included several new sets of experimental data that clearly illustrate the three key findings outlined above. We hope that the new main text and figures communicate our achievements in ways that are now much more accessible to a non-specialist reader.

Lastly, we very much appreciate the referee's commendation on our "*honest and thorough attempt to develop the best classical method*".

Text modifications:

- Figures and main text of the manuscript have been significantly reworked to simplify technical contents and focus on the three main findings related to quantum advantage as outlined above.

Comment: While several classical numerical methods are explored, I am unaware of any theoretical result that shows exponential separation between (noisy) quantum and classical computing on the problems discussed. In fact, following arXiv:2407.12768 it seems plausible that an extension to OTOC's would formally be polynomial albeit very impractical. This does beg the question whether some more refined approximate methods wouldn't outperform those of the authors. Possibly some adaptation of DOI: 10.1126/sciadv.adk4321.

Response: We thank the referee for raising this question regarding numerical simulation. We will first answer the question regarding exponential separation, followed by a response to the various numerical methods the referee raised.

<u>On the quantum (noisy) – classical separation</u>

Our work does not depend on or claim an exponential separation between *noisy* quantum and classical computing, although an exponential separation between *noiseless* quantum and classical computing may indeed exist for the problems discussed. Instead, our paper reports on experimental measurements of *noisy* $\mathrm{OTOC}^{(2)}$ which is highly correlated with the ideal noise-free $\mathrm{OTOC}^{(2)}$. Our conclusion is that this degree of correlation or accuracy is impractical to achieve with currently known classical algorithms running on state-of-the-art supercomputers.

As the referee rightly observes, it is important to investigate whether refinement of any state-of-the-art approximate classical algorithms could outperform our quantum processor. While it is fundamentally impossible to rule out future advancements of classical algorithms, we did attempt to demonstrate that any advancement beyond the algorithms we have presented is quite challenging. This is shown through two means: 1) We theoretically identify that the many-body interference effect in $\mathrm{OTOC}^{(2)}$ is associated with a sign problem and is an essential feature of the system. It presents an intrinsic barrier for classical description of $\mathrm{OTOC}^{(2)}$ and therefore for classical algorithms. 2) Numerically, we have either analyzed or directly checked a large collection of the most performant classical algorithms known to us. None of them has outperformed the quantum processor so far. This is summarized by Table S1 of the SI and the sections referenced therein.

<u>Algorithms suggested by the referee</u>

We thank the referee for pointing out the necessity to cover the Pauli-weight truncation (arXiv:2407.12768, 10.1126/sciadv.adk4321) and the belief-propagation / MPS (10.1126/sciadv.adk4321) methods as potential candidates for the $\mathrm{OTOC}^{(2)}$ simulation. We give theoretical arguments below why we expect failure of these methods. Additionally, following the referee's comment, we also included a report of our implementation and testing of these methods in Supplement.III.C.6-8.

Pauli-weight truncation methods.

Regarding arXiv:2407.12768, that work establishes that as the number of qubits n increases at a fixed noise rate gamma, the complexity of the noisy observable remains at most polynomial in n. However, the degree of the polynomial, approximately n^(1/gamma), can be extraordinarily high. In the present scenario, this bound gives an impractical runtime of approximately n^400 given the current noise rate.

In SI.III.C.6, we show that Pauli path algorithms have a very large cost in our case because only large Hamming weight Paulis contribute non-trivially to OTOC and $\mathrm{OTOC}^{(2)}$. We also run a simulation in a 40-qubit system, which indeed shows little correlation with exact simulation.

MPS-based methods: MPS, MPO and belief propagation.

Another family of methods mentioned in 10.1126/sciadv.adk4321 is based on various tensor-network representations (MPS, MPO, PEPS) of the wave function. Besides, this reference employs belief propagation, a novel method for approximately contracting large 2D tensor networks represented in a form of PEPS.

The tensor-network representations used in reference 10.1126/sciadv.adk4321 are effective at simulating low-entangled states (see https://arxiv.org/pdf/2306.14887, Fig 5, right-bottom panel). Similarly, the more recent applications reported in https://arxiv.org/abs/2503.08247 and https://arxiv.org/abs/2503.05693 focused on near-adiabatic evolution with limited entanglement.

In contrast to these previous works, the analysis we present in SI.III.C.7 (See Fig. S38-S40) demonstrates significant, volume-law, entanglement in our experiments. This implies that the tensor-network representations fail if applied to the OTOC$^{(2)}$ experiment.

Nevertheless, one may wonder if aiming at SNR similar to that in the experiment (instead of a perfect simulation) could significantly lower the required bond dimension. We find this also not to be the case in SI.III.C.7.

Text modifications:

- Expanded the supplement, added Supplements.III.C.6-8.

Comment: To be able to actually estimate the OTOC efficiently on the quantum device, it can not be too small since it's estimated from the average outcome of a single qubit measurement. Slowering scrambling circuits are therefore proposed, but it's not clear to me how this affects the hardness. Naively, it would mean that Pauli strings also spread less quickly in the Pauli space and if this effect is parametrically different it should also make classical simulations more efficient. For example, it's not clear to me that this doesn't make it easier to predict the circuit-to-circuit fluctuations.

Response: It is true that our experiment adopted slower-scrambling circuits (compared to our previous work) in order to optimize the tradeoff between signal size and complexity. The impact on complexity due to the chosen circuit structures is explicitly accounted for in our evaluations of various classical simulation algorithms in the supplement. All tensor networks contraction cost estimates, Monte-Carlo simulations or other heuristic algorithms are based on the exact same circuits run in the experiment.

Also, we note that the slow-scrambling circuits used in our work are in fact more general than the maximally scrambling circuits. As shown by the analytics and numerics in Section IV of the SI, Haar random circuits exhibit wavefront broadening in OTOC and OTOC$^{(2)}$ similar to the circuits used in our current work, whereas the ballistically propagating wavefront is only a special case of 2D circuits ("dual-unitary" circuits). As such, our result is generalizable to other non-integrable dynamics in 2D.

Comment: Error mitigation is used to correct for some effects of the noise. That this creates some uncertainty in the results can be seen from Fig 2b. The OTOC can not exceed 1 if it was estimated without any post processing. It's clear at 15-20 cycles qubits that are still outside the lightcone start to have some fluctuating OTOC which exceeds 1. I consider this as an estimator of the error bars on the mitigation. It's of course hard to estimate from the plot, but it's visible by the naked eye. Turning to panel 2e, it looks like these O(10-3) ought to be more affected by this mitigation procedure. The fact that they are not, suggests that these measurements are more robust, again begging the question whether they can not be computed more easily classically.For example, it's not clear to me that this doesn't make it easier to predict the circuit-to-circuit fluctuations.

Response: We thank the referee for pointing this out. Indeed, our previous manuscript did not sufficiently explain the systematic errors behind the experimental measurements. In the revised manuscript and Section II.F.3 of the supplement, we now provide a model for the systematic errors associated with each OTOC or $\mathrm{OTOC}^{(2)}$ measurement. In summary, the error of each $\mathrm{OTOC}^{(k)}$ constitutes two contributions: A "multiplicative" contribution which scales linearly with the size of the signal, and an additive component independent of signal size. The multiplicative contribution is understood to originate from fidelity fluctuations from circuit to circuit, whereas the additive noise is generally dominated by statistical shot noise coming from a limited number of measurement shots. For large OTOC or $\mathrm{OTOC}^{(2)}$ values, the experimental error is dominated by the multiplicative error which scales with signal size. This explains the fluctuation even when OTOC~1, as pointed out by the referee. For smaller signal sizes, the multiplicative error becomes correspondingly smaller and the total experimental error goes down as well.

The validity of the error model above is vetted by six sets of experimental data ranging from 18 to 40 qubits (Fig. 4d and 4c of the revised main text) and noisy numerics shown in the supplement, from which we provide a bound on the SNR of the 65-qubit data set in Fig. 4a and also circuit-dependent error bars that change according to signal sizes.

Lastly, the finite accuracy of the experiment has been explicitly accounted for when we estimate the complexity of various classical simulation methods. For example, we only estimate the time required to produce one single projection in TN contraction, i.e. the minimum time to obtain any result from the algorithm. For the various Monte-Carlo simulations we attempted, we also use the experimental SNR as the criterion for success (or failure).

Text modifications:

- Fig. 4b and Fig. 4c are now added to illustrate the scaling of experimental error against signal size.

- Discussion in the main text (in "*Quantum advantage with OTOC$^{(2)}$*") now includes a paragraph describing the scaling of experimental error with signal size.
- Error bars on the 40- and 65-qubit circuit fluctuation data in the main text are adjusted according to the error model to reflect larger uncertainty for larger signal sizes.
- Section II.F.3 ("*Residual errors in error-mitigated OTOC$^{(2)}$*") is added in the SI to provide theoretical consideration behind the error model and its validation using experimental data as well as noisy numerics.

Comment: In a previous work Google already reported on OTOC extracted from an earlier device: DOI: 10.1126/science.abg5029. I find that most of the discussion and the philosophy behind these measurements is very similar. So similar, that I am not sure what the added value of this discussion is.

Response: This is an excellent point regarding our presentation. As can be seen in the revised manuscript, we have shifted our focus entirely to the second-order OTOC, i.e. OTOC$^{(2)}$, which manifests the phenomenon of large-loop interference that is crucially absent in the lower-order OTOC$^{(1)}$ studied in the previous work. Furthermore, the polynomial decay of higher orders of OTOC in generic non-integrable dynamics is another new finding of our work and was not realized at the time of our previous OTOC publication. These two findings are the central ingredients behind the quantum advantage demonstration, and the focus of the revised main text discussions.

Text modifications:
- Fig. 2 has been modified to focus on average and fluctuation of OTOC$^{(2)}$, as well as the decay of fluctuations for different orders of OTOC.
- Fig. 3 has been modified to illustrate the mechanism of large-loop interference and how it manifests in OTOC$^{(2)}$.
- Fig. 5 is now added to illustrate how OTOC$^{(2)}$ may be used to infer useful details of quantum dynamics.
- In the main text, the motivation for the experiment is rewritten to highlight the mechanism of large-loop interference in higher-order (i.e. $k >= 2$) OTOCs.
- Most discussions of OTOCs, such as average OTOC and OTOC fluctuation measurements in Fig. 2 and Fig. 3 of the previous version, have been delegated to the supplement.

Comment: I find the discussion of real-world applications, in particular the matching of quantum simulations to experimental data, a bizarre twist. It has barely anything to do with the results presented in the paper and there are 7 pages of supplementary material devoted to it. I look forward to seeing this implemented on Google's device, but since this was not reported in the paper I don't see the rationale.

Response: We agree with the referee that the NMR discussion in the original manuscript seems out of place. It is an ongoing project within our team and we are particularly excited about its eventual realization, but for the present publication it is indeed best to stay close to existing experimental data. In the revised manuscript, we have removed most NMR discussions from the main text, and entirely from the supplement. Instead, we provide a concrete experimental demonstration of Hamiltonian learning (Fig. 5) in the main text, using random circuits as a toy model. Despite the simplicity of the setup, this additional experiment serves as a proof-of-principle illustration on how quantum simulation of $\mathrm{OTOC}^{(2)}$ (or OTOC) can be used to learn useful details of quantum dynamics.

Text modifications:

- Discussion of NMR mostly removed from main and completely removed from SI.
- Fig. 5 added as a concrete example of Hamiltonian learning, along with a short outlook paragraph toward real world applications such as NMR.

Response to Referee #2

Comment: The readout on this processor has significantly improved (along with everything else, mostly due to T1, as the authors state). But what is the readout time? I couldn't find it anywhere. Consider including this, along with median, 1QB and 2QB gate times in the supplement.

Response: We thank the referee for this detailed comment. The gate and readout times have now been added to the supplemental information (Section II.A).

Text modifications:

- The following sentence has been added to the SI: "*The single-qubit gate, two-qubit gate, and readout times are 30 ns, 29 ns, and 600 ns respectively.*"

Comment: Fig 2b: Here (and for fig. 2c) the authors write in caption "with different qubits as Q_b". I would encourage the authors to visualize this in the figure diagrams in some way. It is hard to build an intuition for the data with the vague description "different qubits as Q_b".

Response: We thank the referee for pointing out this sentence. In the revised manuscript, we have modified this sentence in the caption and also rewritten the description of the experiment in the main text that indicates clearly the various dependencies of $C^{(4)}$ and exactly what parameters are being varied.

Text modifications:

- In the main text, we now clearly describe the various parameters determining $C^{(4)}$ and what is being varied when taking the data in Fig. 2: "*Fixing then the number of circuit cycles t within U, the quantity $C^{(2k)}$ (t, q_m, q_b, i) is repeatedly measured… The upper panel of Fig. 2b displays the values of \$\overline{C}$^{(4)}$ (t, q_b), where the overline denotes averaging over circuit instances, for different circuit cycles and choices of q_b. The location of q_m is fixed throughout these measurements…*"
- In the caption for Fig. 2b, we changed the description to "*Color at each qubit site indicates data collected with B applied to the given qubit (Fig. 1b)…*" which refers back to the circuit schematic in Fig. 1b.

Comment: Fig 2c: For certain values (e.g. distance ~24 and cycle ~23) the avg. OTOC is larger than 1. What is the interpretation of this?

Response: Great question! A related question was also raised by Referee 1. As explained in the main text (Fig. 4b discussion), errors in experimental measurements have two contributions: A "multiplicative" contribution that scales with signal size, and an "additive" contribution that is independent of signal size. For large signal sizes (such as 1), the multiplicative contribution completely dominates the error and can be large. For small signal sizes (~0.01), the errors scale down accordingly and are usually a composite of the multiplicative and additive contributions. In Section II.F.3 of the supplement, we provide numerical and experimental evidence for these two error contributions, and theoretically identify the origin of the multiplicative error as fidelity fluctuations between circuit instances.

Text modifications:

- Fig. 4b and Fig. 4c are now added to illustrate the scaling of experimental error against signal size.
- Discussion in the main text (in "*Quantum advantage with OTOC$^{(2)}$*") now includes a paragraph describing the scaling of experimental error with signal size.
- Error bars on the 40- and 65-qubit circuit fluctuation data in the main text are adjusted according to the error model to reflect larger uncertainty for larger signal sizes.
- Section II.F.3 ("*Residual errors in error-mitigated OTOC$^{(2)}$*") is added in the SI to provide theoretical consideration behind the error model and its validation using experimental data as well as noisy numerics.

Comment: Fig 2d: The double-x axis for OTOC(s) and the TOC is confusing. Why are the TOC Cycle axis (top) chosen to go to 19? Why not to 10 where the last datapoint is, like OTOC axis? Perhaps this is to make the gray and purple decays 'similar', but is it perhaps misleading? Not a crucial point, but something to consider.

Response: The referee makes a good point about the confusing double axis. In the revised manuscript, we have updated this plot which is now Fig. 2c. Here, TOC is now along the same axis as OTOC, OTOC$^{(2)}$ and the off-diagonal OTOC$^{(2)}$. Note that the decay for OTOCs is different from our previous version because we have chosen now to measure fluctuation in the middle of the wavefront, i.e. where average OTOC is ~0.5. This allows us to investigate early time decay of OTOCs as well, so as to enable a more "apple-to-apple" comparison with TOC.

Text modifications:

- Fig. 2c (updated version of Fig. 2d in the initial manuscript) now has a single time axis.

Comment: Fig 2d: I would propose to draw a 'TOC' circuit diagram somewhere (just in supplement would be fine).

Response: We have now added the circuit diagram for TOC in the supplement.

Text modifications:

- Circuit diagram for TOC added to Section II.C.1 of the SI.

Comment: Fig 2d: I think an inset showing the experimental setup (indicating Qm and Qb would be really helpful, like in Fig. 2e and f). I couldn't find a description of how many qubits are involved in the data in Fig. 2d)

Response: We thank the referee for pointing this out. The original version of this figure has been updated to Fig. 2c of the current version, where the measurement protocol has changed (as mentioned in the response above). Instead of fixing the choice of q_b, we in fact adjust its location such that our measurement always takes place near the wavefront where average-OTOC is ~0.5. This illustrates the non-exponential decay more clearly and allows experimental results to be compared to theoretical studies conducted in Section IV of the SI. As such, it is difficult to show the schematics for different choices of q_b in the main text. Instead, we show the layouts and the number of qubits involved for different cycles in the SI.

Text modifications:

- In Section II.C of the SI, a figure showing layouts of Fig. 2c and the number of qubits within the lightcone of q_m vs cycles has been added.

Comment: Fig. 2f: Regarding this data I think the authors should consider including a comment/discussion to explain why, even when averaged over paulis, there are still certain datapoints that have significant C^2 (x 10^3). My understanding of this data is that there are three different families of results:

Type 1: Circuits where outcome is dramatically reduced by averaging [e.g. circuit instance ~29, with value approx. -2.5 (/10^3) and circuit instance ~27 with -5.5 (/10^3)].
Type 2: Circuits where outcome is changed partly when averaging [e.g. circuit 46 and 48 where the averaged value is reduced, but has not gone to zero].
Type 3: Circuits where C^2 is enhanced when averaging (e.g. circuit ~7, where yellow point and black sim are at about 5 (/10^3).
In all cases, the impressive MC simulation clearly captures the averaged value (as it should)

- Based on the discussion in the main matter of the paper I find there to not really be a discussion of neither Type 2 nor Type 3 circuits. Perhaps the authors could clarify this? This discussion (again, in the supplement is totally fine) would help the reader navigate these types of plots.

- Is the intuition here that if one was to average enough then all circuits would be of Type 1? Or is there something intrinsic going on here?

Response: We thank the reviewer for this detailed observation. The three different behaviors noted by the reviewer can be understood as follows: Applying the Pauli channel effectively splits

OTOC$^{(1)}$ into two contributions, $C^{(2)} = C_{diag} + C_{loops}$. C_{diag} is a classical contribution reproducible via Monte-Carlo whereas C_{loops} is a quantum contribution coming from small interference loops. C_{diag} is invariant under random Pauli insertion since it is insensitive to the phases of Pauli strings, whereas the phase-sensitive C_{loops} vanishes upon averaging over the Paulis. Moreover, C_{diag} and C_{loops} are circuit-dependent and largely uncorrelated. Therefore the Pauli channel averaging results in a random shift C_{loops} that can either have the same or opposite sign as $C^{(2)}$, leading to either an enhancement or suppression of signal amplitudes of individual circuit instances.

Regarding the last question on whether all circuits will be type 1 upon enough averaging, we believe the answer to be no since Monte-carlo simulation in the figure corresponds to C_{diag} (the remaining quantity after infinite averaging) exactly and is non-zero. The good agreement between Monte-Carlo simulation and experimental results after Pauli-averaging indicates enough random Paulis have been used in experiment that the results have converged. So indeed the remaining fluctuation is intrinsic to the quantum circuits.

Text modifications:

- Fig. 2f of the initial manuscript has now been moved to Section II.E.2 of the supplement. In the same section, we noted the observations raised by the referee and provided a simple explanation: "*In particular, we find that some circuit instances (e. g. i = 6) are enhanced by the averaging process while other instances are suppressed by the averaging process (e. g. 46 and 48). This relative fluctuation may be understood by the fact that the interference contribution to $C^{(2)}$ can have either the same or the opposite sign as $C^{(2)}$, and therefore subtracting its value may lead to either suppression or enhancement of the experimental signal.*"

Comment: I found the discussion surrounding Fig 3c in the main text oddly hard to follow. In particular, is the data in Fig.3c bottom (purple) just the difference between Fig. 3c top (red and blue curves)? If so, then I think standardizing the visual presentation of these datasets and replacing the legend in Fig. 3c bottom with some more explicit would be helpful. A sentence in the main text explaining why Paulis injected at middle of U (versus ~1/3 and ~2/3 through U) does not kill interference would also be helpful. Currently there is just a sentence in the paper saying it is so, but not why.

Response: The referee made a good point about our unclear presentation on interference in the initial manuscript. We have largely rewritten the discussion on interference and combined (new) experimental data for both OTOC and OTOC$^{(2)}$ into a single figure (Fig. 3c of the current manuscript). Here, we compare the impact of inserting random Paulis at different circuit cycles on the resulting OTOC or OTOC$^{(2)}$ values, where we find much larger changes in OTOC$^{(2)}$ due to the presence of large interference loops. We hope that this figure is much easier to read compared to the previous version. The weaker impact of inserting Paulis toward the middle compared to the edges, in the case of $OTOC^{(2)}$, is also explained in the main text as a consequence of gates in the middle U being further from M and B operators and therefore having less impact on the first-order OTOC.

Text modifications:

- When discussing the new experimental data in Fig. 3c, we made the following remark: "*The larger residual effect of Pauli insertion near the edges of U is attributed to the fact that quantum gates closer to the B and M operators have more weights in the resulting signals.*"

Comment: Fig 3c top (minor): Would it make sense to plot this data on a log scale? Just to help the reader.

Response: We thank the referee for this comment. Since the previous version of Fig. 3c has been removed from the main text, this is no longer actionable. In the new Fig. 3c, we hope the way we present the data is much easier to read.

Comment: Fig 3c bottom (minor): I believe the y-axis is mislabeled. Should have been ΔC^((2)) (Delta is missing).

Response: We thank the referee for pointing out this typo. As mentioned above, the previous Fig. 3c has now been removed from the main text.

Comment: On p.4 top left the authors estimate '3 days on Frontier […]'. I think the authors should add somewhere around here the wallclock time to sample the OTOC2 on the quantum processor.

Response: The wallclock time to sample data in Fig. 2f of the initial manuscript is about 30 minutes per circuit. This has now been noted in the SI section where the figure has been moved.

Text modifications:

- Fig. 2f of the initial manuscript has now been moved to Section II.E.2 of the supplement. In the same section, we note: "*Simulating each circuit in Fig. S7 using TN contraction algorithms is estimated to require ~3 days on the Frontier supercomputer, in comparison to ~30 minutes of data acquisition time on the quantum processor.*"

Comment: On p.5: I found the introduction of the Loschmidt echo to be quite ad hoc. I would strongly urge authors to better motivate this choice. Even when looking in the supplement it still begs the question (at least to me): why are we introducing this fidelity metric? It was also used in a related context in a paper also on superconducting qubits and measuring OTOCs (Braumueller et al, Nat Phys 2021, which could be added as a citation). There is of course nothing wrong with the loschmidt echo fidelity, but readers might wonder why it is suddenly introduced here.

Response: Indeed the Loschmidt echo in the initial manuscript was not sufficiently motivated. The choice was primarily because it closely approximates the noise-damping factor on $C^{(4)}_{\text{off-diag}}$, as shown in Fig. S13 of the revised manuscript. Since SNR is closely related to the noise damping factor, using echo allows us to estimate the noise damping factor and therefore the SNR bounds of the 65-qubit dataset. In the revised manuscript, we have provided a more detailed phenomenological error model for $C^{(4)}_{\text{off-diag}}$ and validated it using small-scale experimental benchmarks up to 40 qubits. This allows the SNR of the 65-qubit data set to be estimated without needing to resort to echo. As such, mentions of Loschmidt echoes have been removed from the main text.

We have also cited the Nat Phys paper mentioned by the referee in the main text, as part of the past works on OTOC.

Text modifications:

- Fig. 4e of the initial manuscript has been removed. Instead, we provide an SNR estimate using a more detailed error model summarized by Fig. 4b and Fig. 4c (and related discussion).
- Braumueller *et al*, Nat Phys 2021 added as a reference in the main text.

Comment: Fig 4e (minor): I found the legend hard to read. Perhaps replace (1), (2) and (3) with something like (105q), (72qv1), 72q(v2) or something like that. Obviously a very minor point.

Response: Fig. 4e of the main text has now been removed.

Comment: In VI: Methods – Choice of initial states. Please refer directly to the point in the SI where the 66 qubit circuits which mimic OTOCs are (for other points that refer to SI you refer directly to the section(s) in the supplement that's relevant)

Response: We have added a reference to the specific SI section related to this point.

Comment: Supplementary Fig. S35 (minuscule): Typo in the title.

Response: Typo fixed.

Response to Referee #3

Comment: In order to show any practical quantum advantage, as argued in the manuscript, it is key to go beyond mere academic questions such as sampling from a random circuit. In the present work, OTOCs are measured, which contain more physics, such as about scrambling and thermalization dynamics. However, it is not so clear to me what one can actually learn in terms of physics from OTOCs and its higher-order variant OTOC^2. Although there has been significant theoretical research on OTOCs in quantum matter within the last decade, it has remained open to me for what kind of general physical question it is the OTOCs which could give us new insights (despite of the rather theoretical question of operator spreading). While I understand that in the outlook Hamiltonian learning in the context of NMR is mentioned (and further discussed in the supplementary), it is not clear to me that OTOCs are the key element and whether for instance other measurements could also allow for doing Hamiltonian learning. The manuscript would profit significantly from discussing the physical importance of OTOCs and the higher-order variant OTOC^2.

Response: The referee made several great points about our presentation. Indeed, the previous manuscript version was overloaded with details which obscured the key achievements and findings. Following the suggestion by the referee, we have completely rewritten the manuscript. It now clearly defines the three criteria of useful quantum advantage in the beginning: 1) Large signal size. 2) High classical simulation complexity. 3) Practical value in modeling quantum dynamics. The text then focuses on a set of measurements that illustrate the fundamental physics findings behind each criterion: In Fig. 2, we find that OTOC and its higher order variants exhibit power-law decay in non-integrable dynamics, which is contrasted with the exponential signal decay for ordinary correlators. This satisfies Criterion 1. In Fig. 3, we describe the physics of quantum interference in the space of Pauli operators and show that higher-order OTOCs possess large-loop interference that cannot be classically approximated. This satisfied Criterion 2. In Fig. 4, we demonstrate quantum advantage with a set of large-scale circuit measurements. In Fig. 5, we focus on Criterion 3 and show experimentally measured $OTOC^{(2)}$ values allow parameters of a many-body unitary to be extracted with good accuracy. We hope the logic flow, reworked figures and physics insights contained in the updated manuscript address this comment from the referee.

Regarding Hamiltonian learning, our revised manuscript indeed shows that OTOCs are better suited for this task compared to time-ordered correlation functions measured in most quantum simulation platforms. This is seen by Fig. 2c, where time-ordered correlators decay quickly toward 0 and become featureless, whereas first- and second-order OTOCs decay polynomially and remain sensitive to circuit details at much larger circuit depths. The advantage of OTOCs in learning has also been recognized in published literature (e.g. *Phys. Rev. Research* 5, 043284) due to the same reason. We hope that the updated Fig. 2c and the explicit

Hamiltonian learning demonstration of Fig. 5 have now made the application value of our work clear.

Lastly, we acknowledge that while our work does show a proof-of-principle demonstration of Hamiltonian learning on a toy model, we have not applied this scheme to a "real-world" quantum system such as NMR yet. As such, we have removed the word "practical" from the title and other prominent places of the manuscript.

Text modifications:

- Main text of the manuscript has been extensively rewritten to highlight physics findings.
- Fig. 5 has been added to show a toy-model Hamiltonian learning experiment.
- The word "practical" has been removed from the title, which now highlights many-body quantum interference instead.

Comment: What I find particularly challenging to clearly assess positively is the claim about quantum advantage. It is very much appreciated that the present work puts a lot of effort in comparing the experimental results to outcomes of potential classical algorithms. When it comes to the numerically exact simulation of the circuits, the authors, however, take as the standard reference tensor network methods, and for concrete comparison one particular variant. It is not at all clear to me that this naturally exhausts potential candidates in theory and justifies the claim of (practical) quantum advantage.

While a lot of research in theory targets to solve 2D quantum dynamics with tensor networks, it is, however, not at all clear according to the current state of the art, whether tensor networks are the ideal reference candidate (as it is taken in the present manuscript). In fact, it is generally accepted that tensor networks will necessarily fail to describe thermalization dynamics in 2D. In this context a very promising alternative method, which for dynamics in 2D has already demonstrated to be often superior compared to tensor networks, is variational Monte Carlo, or in more general terms neural quantum states. Because of the recent advances of these Monte-Carlo inspired algorithms, I would suggest to revise the manuscript to properly account for this class of methods, when claiming quantum advantage.

Response: In our new manuscript, we argue that there are no state-of-the-art approximate classical algorithms that could outperform our quantum processor. We approach this in two ways: (1) we identify a many-body interference effect in $\mathrm{OTOC}^{(2)}$ that is associated with a sign problem and is an essential feature of the system. It presents a barrier for classical description of $\mathrm{OTOC}^{(2)}$ and therefore for classical algorithms. (2) We analyze or explicitly check a number of the most performant classical algorithms for computing $\mathrm{OTOC}^{(2)}$, and find that none outperform the quantum processor, see Table S1 in the SI and the sections referenced therein.

We thank the referee for pointing out the necessity to address variational methods. For a case of a genetic circuit, neural quantum states (NQS) is a natural choice due to its flexibility. To assess the applicability of this method, we implement a NQS supervised learning simulation and report the results in an additional supplement section Supplement.III.C.8.

We would like to first give a general argument why we believe this method would fail. Recent successful applications of NQS to time evolution (https://arxiv.org/abs/2503.08247v1, https://arxiv.org/pdf/2403.05147) address either adiabatic or short-time evolution, where the entanglement (or energy density) is low.

As shown in Supplement III.C.7, our time-evolution in the case of beyond-classical dataset is volume law entangled. For such systems, NQS is generally expected to require exponentially many parameters to succeed [https://arxiv.org/pdf/2212.02204, https://arxiv.org/abs/2104.10696].

We directly test this logic in Supplement.III.C.8. We observe that fidelity over 0.5 (loosely, giving SNR > 1) requires a parameter count of order of the Hilbert space size, in full accord with our intuition about representing a near-volume-law state. Training of such a neural network would therefore be exponentially costly.

Text modifications:

- Expanded the supplement, added Supplements.III.C.6-8.

Comment: Large parts of the manuscript aim at demonstrating the importance of quantum mechanical interference in the dynamics of the OTOCs. This is important because many aspects of thermalization dynamics can be actually described purely classically in terms of hydrodynamic equations of motion (this could be also mentioned in the main text as a limitation of claiming quantum advantage). I have to admit that the presentation in the manuscript about the precise setup and implications of the random injections of Pauli operators is not very clear to me. When comparing the data of the original OTOC and OTOC^2 to the ones with random Pauli injections, one can certainly see some quantitative changes in the results. However, it also appeared to me that there is no qualitative change. In order to actually settle whether quantum interference is of key importance, it would be necessary to study this systematically as a function of system size. Overall, it is not clear to me of how fundamental importance quantum interference actually is here, which, however, might also be rooted in not properly appreciating the random injections due to an unclear presentation.

Response: The referee raised excellent points regarding our interference experiments. The revised manuscript now discusses this effect in a more focused manner, and provides a clear contrast between OTOC and OTOC$^{(2)}$. Here, let us also specifically answer the referee's question regarding "key importance" of interference and whether this is only a transient effect that disappears at large system sizes.

As explained now in the revised manuscript, there are two types of quantum interference: "Small-loop" interference that may be visualized by space-time loops formed during the time-evolution of B(t), and "large-loop" interference coming from the fact that Pauli strings contributing to experimental observables at the final time can in general form a loop with arbitrarily large area. The small-loop interference effects are much weaker and our figure of merit, 1-correlation, is indeed expected to disappear for large system sizes. This is clearly illustrated by the effect of a single random Pauli layer on $C^{(2)}$ (which only has small-loop interference) shown in Fig. 3c of the updated manuscript: As the location of the random Paulis is shifted into the middle of U, the Paulis almost cease to have an effect on $C^{(2)}$. This experiment indicates that only Paulis near the B or M operators have a large impact on $C^{(2)}$, and therefore as the size of U grows bigger, the overall impact of interference may decrease.

The above situation is quite different for large-loop interference, which is present in $C^{(4)}$. This is also evident from the data shown in Fig. 3c: First, unlike $C^{(2)}$, random Paulis produce a much more dramatic change in the $C^{(4)}$ as seen by the very poor correlation between the Pauli-averaged and non-averaged signals. Second, the strong spatial dependence of the signal change seen in $C^{(2)}$ is absent in $C^{(4)}$; the correlation remains poor regardless of the cycle at which Paulis are inserted. This leads us to conclude that large-loop interference is a substantial contribution to $C^{(4)}$ and is not localized in spacetime as the small-loop interference is. It is therefore not expected to disappear at large system sizes.

To further solidify the points made above, we have included a new figure in Section II.F.5 of the SI which shows the signal change as a result of Pauli insertion into the middle of U, for system sizes from 18 to 40. It can be seen that the signal change, characterized by 1 - correlation, barely decays over system sizes for $C^{(4)}$. Even the small residual decay is attributed to effects of noise after comparing with noiseless simulation. This result, along with Fig. 3c of the main text, is sufficient to indicate that large-loop interference remains a significant contribution to $C^{(4)}$ even as system size grows.

Text modifications:

- The section on quantum interference has been rewritten with new experimental data to clearly illustrate the underlying physics.
- Fig. 3c has been modified and Section II.F.5 added to the SI to show that large-loop interference contribution to $C^{(4)}$ does not decay as system size increases.

Comment: In the outlook the manuscript speculates that the hardware development "will rapidly increase the margin of quantum advantage, outrunning any gains that may result from improving classical simulation algorithms". This speculative comment might be correct when taking as a reference tensor networks, but it might not be for alternative numerical methods, which don't suffer from entanglement growth or matrix contraction complexity. I would suggest to adapt this speculation accordingly.

Response: We agree with the referee that the language has to be more precise. Indeed exponential growth of complexity generally applies to methods such as MPS and tensor contraction, and heuristic algorithms like the Monte-Carlo simulation we considered in the supplement do not necessarily suffer from the same scaling. On the other hand, since all of the heuristic algorithms we have tried also failed against OTOC$^{(2)}$, the only viable simulation methods are ones that have exponentially growing complexity.

Nevertheless, we do acknowledge that our work cannot rule out any future heuristic algorithm that is more efficient than the ones we have considered, so we have adapted this closing statement to more precisely refer to tensor contraction, as the referee recommended.

Text modifications:

- The sentence in the conclusion has been modified to read "*...continual improvement of quantum processors will also widen this gap by exponentially increasing the cost of TN contraction and allowing for the realization of even more classically intractable interference phenomena.*"

Comment: On page 2 it is speculated that due to the change of the spectrum of the correlation operator a phase transition is expected. Is this meant to be a phase transition in time or as a function of some other control parameter? Could it also be a (sharp) crossover instead?

Response: We thank the referee for pointing this out. The phase transition in the spectrum of the correlation operator B(t)M is a dynamical transition that occurs as a function of time t. Specifically, as t crosses its critical value t*, the spectrum turns from being gapped to being gapless.

In the thermodynamic limit, an initially gapped spectrum cannot turn gapless immediately with the increase of time from t = 0. In this case, we expect the evolution across t* to be a phase transition rather than a crossover. We verify this expectation explicitly in a random matrix theory-based toy model of the B(t)M spectrum, see Section IV.B of the SI.

The referee is correct that, in a finite size system, the crisp phase transition smears into the crossover. However, in the present context, the sharpness of the smeared transition is controlled by the inverse Hilbert space dimension scaling exponentially with the number of qubits. This means that the phase transition picture works well even in systems of moderate size (e.g., for 12 qubits), see Figs. S50 and S52.

In the revised manuscript, to reduce the amount of theoretical content in the main text in order to improve accessibility, we have delegated discussions of this phase transition entirely to Section IV.B of the SI.

Text modifications:

- Phase transition theoretical discussions have been removed from main text and fully delegated to the SI.

Comment: On page 2 it is said that a particular classical simulation would require 10^25 years. To me it is not clear whether this number for this one chosen reference candidate can be seen as representative.

Response: We thank the referee for this question regarding RCS. This estimate is based on the complexity of tensor network contraction for the computation of the probability amplitude of a single output bitstring. The computation of probability amplitudes of output bitstrings through tensor network contraction is the primitive of *all* successful simulation methods of the random circuit sampling (RCS) experiment of 2019 (including the celebrated PRL 129.9 (2022): 090502, arXiv:2108.05665, arXiv:2112.15083, and arXiv:2110.14502). Furthermore, tensor network contraction achieves the lowest simulation time estimates for the subsequent RCS experiments from USTC, Quantinuum, and Google.

There is however a phase transition in the complexity of RCS when gate errors are above some threshold (Nature 634.8033 (2024): 328-333). And there are indeed methods that exploit this, making the simulation of RCS accessible under the right circumstances (see PRX Quantum 5.1 (2024): 010334, PRX 10.4 (2020): 041038, PRX Quantum 4.2 (2023): 020304 for representative examples). However, as studied in detail in Google's recent RCS paper (Nature 634.8033 (2024): 328-333), Google's RCS experiments (including the current) have gate errors much lower than this threshold, and are therefore protected against this kind of attacks. This leaves the optimization of tensor network contraction of the RCS circuits as the best known classical method for attempting the simulation.

In addition, the cost of tensor network contraction depends heavily on the memory available by the classical computer performing the computation. The unrealistic assumption of having no memory bounds provides the lowest cost possible. Reducing the memory available increases the cost exponentially. In the present work we assume no memory bounds, hence giving only a *lower bound* to the computational cost of tensor network contraction.

Finally, tensor network contraction optimizers have become extremely refined in the last 5 years, and our own optimizer (TNCO) is shown in the current paper to find lower costs than other available optimizers (set II.B, III.A, and III.C.4 of the Supplementary Information).

In conclusion, we provide extensive evidence that the estimate of 10^25 years for the computation of probability amplitudes from the output of the random circuits run in the current work, and hence the task of sampling from them, is a generous lower bound to the cost of all currently known simulation methods.

Comment: I have checked the code, that was provided to me on the portal. I have downloaded, read the instructions, and executed the program. Everything worked out. It is worthwhile to mention that the provided data only refers to Fig. 4f and thus only provides a potential check for this particular figure.

Response: When submitting the revised manuscript, we have included additional small scale circuits for OTOC$^{(2)}$ as well as OTOC$^{(1)}$ so their results can be more easily checked.

Text modifications:

- Circuits used in Fig. 3d, Fig. 3e, Fig. 4a, Fig. 4b and Fig. 4c are all provided. In addition, OTOC ($C^{(2)}$) circuits and data with 95, 40 and 31 qubits are also provided. For system sizes <= 40 qubits, we include exact simulation results. For system sizes > 40, we include experimental data.

Other manuscript changes

- The follow authors have been added:
    - Tom Westerhout (tom.westerhout@ru.nl)
    - Andreas Kabel (aka@google.com)
    - James M. Manyika (jmanyika@google.com)
    - Yossi Matias (yossi@google.com)
    - Elica Kyoseva (ekyoseva@nvidia.com)
    - Ikko Hamamura (ihamamura@nvidia.com)
    - Saul Cohen (scohen@nvidia.com)
    - Hanfeng Gu (henrygu@nvidia.com)

Dear Editor,

Thank you for considering our manuscript originally titled "*Experimental quantum advantage with an observable governed by many-body interference*" and now titled "*Constructive interference at the edge of quantum ergodic dynamics*". We are pleased to see that all three referees have commented on the improved clarity of the manuscript, and the technical concerns raised in the first round of reviews have been addressed. The main feedback raised by Referees 1 and 4 in this round of review pertains largely to the putative nature of our quantum advantage claim.

We would like to first emphasize that we fully rebut all specific comments by Referees 1 and 4 and respond to the questions about quantum advantage, which are highly speculative in nature. Following editorial guidance, we have also reworked our manuscript to fully focus on the physics finding of constructive many-body interference revealed through OTOCs and its scientific usefulness in restoring sensitivity to quantum dynamics. Practical quantum advantage is now only mentioned in the conclusion and outlook sections as an eventual goal (to which our work remains a crucial step toward). If there is any desire to invite other reviewers, we suggest the following individuals as potential referees (excluding past and current collaborators) :

Shinsei Ryu, Princeton University
Titus Neupert, University of Zurich
Brian Swingle, University of Maryland
Romain Vasseur, University of Massachusetts Amherst
Tibor Rakovszky, Budapest University of Technology and Economics
Curt von Keyserlingk, King's College London
John Chalker, Oxford
Jonathan Ruhman, Bar-Ilan university
Michael Gullans, University of Maryland
Brian Skinner, Ohio State

We believe that the newly updated manuscript is fully suitable for publication in Nature and hope that the work may be published without further review.

With our best regards,
The corresponding authors

Response to Referee #1

Comment: The paper presents an experiment on a 103-qubit superconducting quantum processor. The experiment consists of measuring a generalization of the out-of-time-order correlator to a multi-round setup. The results focus on the second-order version in which the echo sequence is repeated twice.

The paper has been substantially revised as compared to the previous version, and many of the concerns have been resolved. The results are more focussed and the presentation is more clear. The main claim of the authors is that "interference between Pauli strings that form large loops in configuration space" makes the classical simulation intractable. That such loops contribute is evidenced by checking the sensitivity of the results to a Pauli-insertion scheme which randomizes the signs of these loops, thus suppressing their contributions.

In the introduction the authors present very well the 3 criteria for a worthwhile demonstration of quantum advantage. I summarize, (i) identify observable with a signal-to-noise ratio that is large enough such that it enables efficient extraction on the noisy quantum device, (ii) the same observable must be sufficiently complex such that it eludes classical simulation, (iii) there needs to be some practical use for the observable.

In my opinion the paper falls flat of (iii), but that's a high bar so let's leave that out for now. My main concern is that there is a fundamental tension between (i) and (ii) that, in spite of the thorough analysis in the supplement of the paper, I do not see resolved. The fact that the otoc only decays algebraically, allows to satisfy (i) but possibly opens the door to classical simulation as the amount of information one needs to capture about the state of the environment is now fundamentally limited. Of course it might well be that no classical algorithm can efficiently find this optimal compression, but that remains to be seen. This leaves the door open for future algorithms to simulate the experiment.

Response: We appreciate that the referee now finds the updated manuscript is more focused and the presentation clearer. However, we respectfully disagree with the referee's assessments in the last paragraph:

- The referee opines that "*the paper falls flat of (iii)*". In Fig. 5 of the updated manuscript, we explicitly showed a Hamiltonian learning example where $\text{OTOC}^{(2)}$ data is used to learn an unknown two-qubit gate within the random circuit. The scheme is readily applicable to practically relevant physical systems, such as those of solid-state NMR. The usefulness of OTOCs in Hamiltonian learning has also been recognized in literature [e.g. Phys. Rev. Research 5, 043284 or https://arxiv.org/abs/2411.12794]. However, we do acknowledge Hamiltonian learning of an actual physical system has not been performed in our work (and it is beyond the scope of this paper). We have therefore further de-emphasized the quantum advantage claim in the revised manuscript (see below).

- The referee raises the point that there is fundamental tension of "algebraic decay" of observables and classical simulatability and on that basis, speculates that our results may be easily simulated. Indeed, this has been identified in the literature as a significant challenge for quantum advantage in computing *expectation values of physical observables*, see for example arXiv:2407.12768. Nonetheless, it has been rigorously proven that computing expectation values is BQP-complete, in other words there exist example circuits with quantum advantage. These most likely will not be generic and instead will have a certain structure. Our second order OTOC circuits introduce a minimal structure while remaining sufficiently generic that circumvents this challenge by exploiting the exponential advantage provided by *including time inversion in the measurement protocol*, see arXiv:2407.07754. As evidenced in the manuscript, time inversion reveals a (large-loop) interference that serves as the physical mechanism underlying properties (i) and (ii) of our circuits. This is by no means a rigorous proof of quantum advantage. Rigorous proofs of quantum advantage, apart from a handful of exceptions, apply to worst case instances (i.e. require fine-tuned circuits). Note that even in the case of Schor's algorithm there is no proof of classical complexity. Instead our goal in this manuscript is to address a computation as close to the practical (typical instance) setting as possible. In this case a rigorous proof would be a major breakthrough beyond the scope of our manuscript. Nonetheless, we were able to identify an inherent challenge analogous to the fermionic sign problem obstructing efficient classical representation (compressed storage) of OTOCs. We demonstrated quantitatively that this sign problem is severe (results in at least exponentially divergent variance of Monte Carlo algorithms and bond dimension of MPO/MPS representations) and inherent to our circuits (persists even in the basis reflecting all local symmetries of our circuits). This physical obstruction to efficient classical algorithms in conjunction with our direct implementation of a whole laundry list of standard and bespoke algorithms allow us to conclude that tensor network contraction is the best performing classical algorithm. Moreover, improving significantly on our estimate of the cost of tensor network contraction algorithm requires a qualitatively new classical algorithm if at all possible.

Recognizing the referee's reservations, we have now modified the manuscript again to fully focus on the new physics of constructive many-body interference observed through our work and its implications on sensitivity of quantum observables. Quantum advantage is mostly mentioned toward the end as a future direction to which our work remains an important step toward. We believe the phrasing of the updated manuscript should address the referee's concerns.

Response to Referee #3

Comment: In this revised version of the paper from the Google Quantum AI team and their collaborators, they have made many deep changes to the paper. While the fundamentals of the original data is still present, several new important (and impressive and clarifying) datasets/analysis are also presented, overall dramatically improving the quality of the manuscript.

While the fundamental dataset related to 'experimental quantum advantage' (essentially, Fig.4) is not changed, the exposition and the arguments leading up to this dataset are now vastly clearer and the impact of the result is significantly stronger. The authors have most certainly taken into consideration both mine (and other referees) comments in these wide-ranging revisions. Overall, this overhaul of the paper dramatically increases its impact in my view: instead of working hard to phrase their result in terms of 'practical' quantum advantage, it is now a study of 'experimental' quantum advantage and on balance the presentation is now much clearer, and the impressiveness stands out better. In particular the argument around interference effects in OTOC(n) values is much more pedagogical (Fig3 a,b,c) and I think will help other researchers better appreciate the physics being used (and perhaps spur others to utilize this technique more, now that it's been clearly laid out). The impact of what is achieved is thus more apparent and easier to understand.

I applaud the authors for now carefully spelling out what they deem a 'current generation quantum advantage claim' must satisfy (I agree with their arguments here – others could reasonably disagree, but it's a quantitative attempt at defining something elusive). In this revised version of the manuscript, they then put their OTOC(n) idea and hardware to the test on this 'benchmark'. While details of when quantum advantage has been definitively achieved are a gray zone (as the authors are well aware and mention in the paper), this paper sets a benchmark for how the community could/should think about quantum advantage in current/near term processors. With the publication of this manuscript in a foremost journal such as Nature I think this approach to claiming 'quantum advantage' will (and should!) become an industry standard – a throwing down of the gauntlet for other teams in the 'current generation quantum advantage' game. The added explicit "definition" of a Goldilocks zone (Fig. 1a) I'm sure we'll be used as a helpful guide for other teams in this community as well.

So, I'm very happy to recommend this manuscript for publication in Nature in its current form. It is an impressive revision which better presents the impressive experiment and offers important insights on how/when the community can claim quantum advantage (beyond random circuit sampling style arguments).

Response: We appreciate the referee's highly positive feedback and support for publication!

Response to Referee #4

Comment: As the authors have adapted the definition of what they claim as quantum advantage, the critique formulated in my first report cannot be directly applied anymore. While the new definition appears reasonable, it certainly matches more closely the results presented in the manuscript. However, criterion iii) on revealing pratically relevant information is only considered theoretically and not demonstrated on the actual quantum device, the reason for which remains unclear. Taking the definition for experimental quantum advantage formulated in the manuscript literally, I don't see their experiment fulfilling the claim.

Response: In Fig. 5 of the updated manuscript, we explicitly showed a Hamiltonian learning example where OTOC$^{(2)}$ data is used to learn an unknown two-qubit gate within the random circuit. This is a proof-of-principle experiment and the scheme is readily applicable to practically relevant physical systems, such as solid-state NMR. The usefulness of OTOCs in Hamiltonian learning has also been recognized in literature [e.g. Phys. Rev. Research 5, 043284 or https://arxiv.org/abs/2411.12794]. The referee is right in pointing out that a learning experiment comparing quantum data with data from an actual physical system has not been conducted by our work. As such, we have modified the manuscript to emphasize the physics findings of constructive many-body interference and sensitivities of OTOCs, which are crucial features that make such observables candidates of practical quantum advantage. The actual demonstration of this advantage using realistic physical systems is left as future works.

Comment: The authors have put forward substantial efforts to further strengthen their claim that the quantum dynamics studied on the quantum processor cannot be simulated efficiently at least for large system sizes. I understand that the key argument is that the classical algorithm would have to learn the intricate sign structure in the quantum states. It is then argued along various axes why the neural quantum states method cannot achieve this.

Response: We thank the reviewer for the specific suggestions to expand our analysis of the NQS application to the OTOC circuits. We significantly expanded our analysis by

- considering both other (i) optimizers and (ii) architectures, including other versions of gradient descents (sgd, sgd with nesterov, cyclic learning rate) which are believed to be less prone to getting stuck in local minima, fully-connected architectures with more hidden layers, convolutional neural networks and networks with complex parametrizations (without the amplitude-phase separation);
- analyzing a typical distribution of fidelities of a circuit instance, generated by starting training from different initial parameters;

- applying NQS to the Pauli representation of the time-evolved B(t) operator;
- developing a systematic study linking NQS success to low wave function of operator entanglement, and showing that due to the large entanglement entropy of our wave functions, an efficient polynomial compression can not be found;
- demonstrating that NQS fidelity is correlated with SNR of OTOC/OTOC2 circuits and therefore is a good predictor of the success of NQS for these circuits.

In our review of the literature, the successful applications of NQS are either in low-entangling setups, or in the cases where there exists an efficient tensor-network representation of a quantum state (with either NQS or MPS). However, the wave functions obtained in the OTOC circuits are neither of those. They are chaotic and close to maximum entanglement (with Renyi entropy over 90%). OTOC circuits also generate near maximal operator entanglement (OTOC magic https://doi.org/10.1073/pnas.2217031120 (2023)), such that neither MPS nor MPO methods provide efficient classical representation of these circuits. We therefore expect NQS to require exponential resources.

Below, we address the particular points of the reviewer:

Comment: First let me point out that neural quantum states are now very successful in learning complex sign structures in frustrated quantum magnets and interacting fermion matter:

https://arxiv.org/abs/2311.16889

https://doi.org/10.1038/s41567-024-02566-1

https://arxiv.org/abs/2405.04472

Response: We agree that when learning the ground state of a quantum system, which is low-entangled (in both gapped and gapless cases, except for exotic models, which are not considered in these references), a neural network can find an efficient low-parameter representation.

Comment: To support their claim the authors included also some analysis in the supplementary material, which seems to indicate that the numerical effort of the neural quantum states method would be exponential in the number of qubits. However, I have quite significant doubts whether the presented results are actually meaningful. First, the performed supervised learning procedure uses adam as the optimizer, which can be stuck in local minima and which it is typically not utilized.

Response: Indeed, in some optimization problems in computer science adaptive methods are believed to fall into "sharp" valleys and may get stuck in local minima (https://arxiv.org/abs/1609.04747, https://arxiv.org/html/2410.10056v1). While no optimizer guarantees convergence to the global minimum, we explore a different optimizer, SGD optimizer with the Nesterov momentum, which is believed to settle into flatter minima (https://arxiv.org/html/2410.10056v1, https://arxiv.org/pdf/1706.02677), with and without the cyclic learning rate (which also helps to avoid local minima) and compare the results against the aforementioned Adam. The results show that the outcome only weakly depends on the optimization scheme. We show the respective plot in the supplement. We note that one of the references mentioned below by the referee (https://arxiv.org/pdf/2412.11778) uses Adam as their main method to obtain results.

Finally, we consider the distribution of outcomes (fidelities) in multiple optimization attempts starting from different initialization of the network parameters. The distribution yields a very low variance-to-mean ratio with no outliers, suggesting that the local minima would not qualitatively change the conclusions of this section.

Comment: Second, the artificial neural network output is split to yield separately magnitude and phase. It is by now well known that such a choice is not suitable to detect nontrivial sign structures. Instead, transformer or CNNs for the full amplitude are typically used. Therefore, it is clear that the author's analysis is not suitable to exclude the neural quantum states method.

Response: We agree that fitting a wave function with complex entries (the notion of "sign structure" is typically used in the tasks of finding the ground state using a NN, as the ground state in many cases can be chosen real-valued) allows for two approaches: (1) either keep the neural network parameters real, and then somehow transform them into the real and imaginary part, which is essentially what we did in the last dense layer of our NN (before that, the NN is joint), or (2) work directly with complex weights and biases and a single complex-valued output representing a wave function entry.

On top of that, there is the question of choice of a particular architecture. In this sense, both real-valued and complex-valued choices (1) and (2) are compatible with choosing a CNN, or fully-connected neural network. For instance, the references [https://arxiv.org/pdf/2002.04613, https://arxiv.org/pdf/2308.09664] use CNN in combination with the amplitude-phase separation. Below, we would like to expand our discussion of CNN and use complex network parameters:

- In the OTOC circuits, the choice of CNN over a fully-connected network is not beneficial given the problem's properties. In condensed-matter systems, CNNs are used to leverage translational symmetry. However, in this work, the chip has no translational symmetry. Most importantly, the random circuit lacks it in general. Our attempt to apply CNN to such a problem leads to a comparable [to the fully-connected architecture] parameter count, required to represent the wave function. In the revised version of the manuscript, we comment on our attempt to apply CNN, show the numerical result and provide rationale for our choice of the architecture. Moreover, we use CNN with complex parameters to avoid any amplitude-phase separation.

- We implemented a version of the fully-connected network, which uses complex biases and weights. This allows us to predict a single wave function entry, without any amplitude-phase separation. Our results yield a very similar Fidelity(number_parameters) curve, showing no practical difference between those choices. We show the respective plot in the supplement.
- Besides, we compare simulations with a fully-connected network and several hidden layers (2, 3, 4). We observe that all Fidelity(number_parameters) curves collapse and show independence on the choice of the neural network depth.

Comment: Further, the references mentioned in the author's report on previous works have left out some of the most important contributions of nonequilibrium dynamics solved with neural quantum states:

https://doi.org/10.1103/PhysRevLett.125.100503

https://doi.org/10.1126/sciadv.abl6850

https://arxiv.org/abs/2412.05332

https://arxiv.org/abs/2412.11778

for up to 400 qubits. Let me note that these examples contain far-from-equilibrium dynamics far beyond what is manageable using tensor networks due to significant entanglement generation.

Response: In the revised version of the manuscript, we cited and commented on all these references. We believe that the successful application of NQS in those particular cases does not imply success in [highly-scrambling] OTOC circuits, as none of these references consider simulation of a state with (a) entanglement reaching volume law values and (b) non-representable by tensor networks. In order:

1) https://arxiv.org/pdf/2412.05332 – we failed to find a mention of time-evolution in this reference;

2) https://www.science.org/doi/10.1126/sciadv.abl6850 – this work studies an adiabatic phase transition. The NQS is used to probe periodic boundary conditions, in which case one can use CNN due to translational invariance. Notably, Fig 3 shows good agreement between MPS- and NQS-obtained results, which signals low entanglement. Despite a considerable number (400 qubits), this system is simulable due to the character of the time-evolution in question. We comment on this reference in the revised manuscript;
3) https://arxiv.org/pdf/2412.11778 and https://doi.org/10.1103/PhysRevLett.125.100503 – those works study the transverse-field Ising model in three regimes, $h / h_c = 2.0$, 1.0 and 0.1. While the first two regimes are low-entangling (with Renyi entropy lower than 20% of the maximum), the last case is near-classical and has an almost-exact D=2 PEPS representation. We comment on those references (and all similar references studying those three points in the TFIM phase diagram) in the revised manuscript;

Comment: Let me emphasize that I consider the thorough analysis of an alternative method such as neural quantum states only important because of the strong claim of practical quantum advantage, which, as I think, makes it necessary to rule out any classical competitor.

Response: We found that state-of-the art NQS techniques do not provide significant reduction in classical computation cost for specific OTOC2 circuits realized in our manuscript. We also identify high state and operator entanglement as a significant challenge for such algorithms. A significant and novel improvement of NQS algorithms for the case of OTOC2 circuit would be required to overcome this. We conclude therefore that tensor network contraction remains the best performing classical algorithm in our case. Moreover, in the revised manuscript we focus the discussion on the observation of the underlying large-loop quantum interference instead of emphasizing the resource requirements for quantum and classical processors. In this way our main conclusions are not relying on specific estimates of efficiency of NQS algorithms.

Comment: In the revised manuscript the authors have significantly improved the presentation on the random injections, which is now much clearer. I very much appreciate the underlying idea to identify quantum interference effects in the dynamics. The new data in Fig. 3c now displays more convincingly that the random injections have a more profound impact of C^4 as compared to C^2. In order to make the presentation in this context finally compelling, it would be important to more quantitatively assess the analysis of the C^4. In the inset of Fig. 3c a fit is presented to extract what is called the Pearson correlation. From the data for C^4 this fit doesn't appear too meaningful due to the wide scattering of points. At least I'd suggest to include the error bars in Fig. 3c which emerge from this fitting. I could imagine that with error bars the difference between C^4 and C^2 might not be that significant anymore.

Response: We have now included error bars in the Pearson correlation coefficients shown in Fig. 3c, obtained by resampling from existing data. The separation between the curves remains. This is perhaps not surprising, given the insets clearly show that $\mathrm{OTOC}^{(1)}$ is perfectly correlated after Pauli insertion whereas $\mathrm{OTOC}^{(2)}$ is barely correlated. The qualitative difference can hardly be accidental.

Comment: I understand that the estimation of classical resources are based on reasonable assumptions. However, what I find difficult to assess in a quantitative fashion is to extrapolate these estimates in order to gauge what might be possible in the near future. For instance, if GPUs could be utilized more efficiently, many orders of magnitude can be easily gained in any of such estimates.

Response: To be as much as agnostic as possible, we have also provided estimates of the computational costs for our large-scale quantum circuits in terms of FLOP (that is, in terms of the total number of floating-point operations). When we provide a "time" for a given architecture, we assume a 20% efficiency of the maximum FLOPs (that is, floating-point per second) that the target classical device provides. In this sense, there are three ways to improve the classical simulations in the future: i) use a better GPU with higher FLOPs, ii) improve the utilization of the GPU (at the moment, we are assuming a 20% utilization, which is typical in large tensor network contraction benchmarks over the last few years), and iii) identify a better contraction path with a smaller FLOP count. However, while we expect i) and ii) to improve in the future, we developed a state-of-the-art tensor network contraction optimizer which outperforms the best existing optimizer to iii) find the close-to-optimal FLOP count for the considered circuits. Moreover, improving i) can only lead to a polynomial speedup, and it is relatively limited for single GPUs (for instance, there is only a ~3x improvement in FLOPs between the 2020 A100 GPU and the 2024 H200 GPU), and improvements in ii) are upper bounded by a factor of 5x.

Dear Editor,

Thank you for considering our manuscript "*Constructive interference at the edge of quantum ergodic dynamics*". We are delighted to see that this version of the manuscript has been positively received by Referee 4, and appreciate the helpful editorial suggestions on final edits of the paper. Below, we provide responses and summarize manuscript changes related to the latest comments by Referee 4 as well as editorial suggestions.

With our best regards,
The corresponding authors

Response to Referee #4

Comment: The manuscript has seen substantial modifications, which is very much appreciated. The restructuring of the presentation away from the claim of quantum advantage towards more addressing physical questions has led to a strong improvement. Overall, the revised version is now clear and accessible.

The experimental results I still consider truly impressive, being capable of experimentally accessing such subtle quantum interference effects in a large many-body system.

Thus, I can now strongly recommend the main manuscript for publication in Nature.

Concerning the authors' response and some parts of the supplement, however, I respectfully disagree.

The authors still seem to not properly present the existing literature and knowledge gained in recent years on neural quantum states, which is important insofar that any claim about advantages over classical algorithms is directly affected by this.

First of all, it is crucial to accept that quantum entanglement, which is critical for most tensor network approaches, is not a limiting factor for neural quantum states, see for instance:

https://doi.org/10.1103/PhysRevX.7.021021
https://doi.org/10.1103/PhysRevLett.122.065301
https://doi.org/10.1103/physrevb.106.205136

Thus, any statement setting a potential failure of neural quantum states into the context of entanglement, as done in the response and the supplement, doesn't appear viable (while I would like to apologize for providing an incorrect reference to time evolution in my previous report).

This becomes slightly more severe in view of the presentation of achieved advances in time evolution for neural quantum states in the supplement, where it is suggested that neural quantum states might be standing behind tensor networks. Actually, the considered works demonstrate clearly (also partly in direct comparison) that neural quantum states can go significantly beyond.

Response: We thank the referee for pointing out the inaccuracy of our original statement. Indeed, there exist cases in which NQS may outperform tensor networks methods. These cases are now cited in the supplement, the most notable examples of which include the volume-law entangled ground states of the quantum Serrington-Kirkpatrick and disordered SYK-type fermionic models https://arxiv.org/pdf/2309.11534.

For generic quantum-chaotic systems, in practice, it is to our knowledge hard to find a polynomial NN architecture that would represent a particular family of volume-law entangled states (such as our OTOC^2 circuits). For example, https://arxiv.org/pdf/2412.11830 does not indicate such sub-exponential scaling. For chaotic wave functions, exponential scaling instead seems unavoidable. Indeed, in the case of unstructured random circuits an exponential number of variational parameters is required for efficient trainability, as was formally demonstrated in https://arxiv.org/abs/2109.06957. In geometries with finite connectivity this effect manifests as either barren plateaus https://arxiv.org/abs/2205.05786 or proliferation of local minima https://arxiv.org/pdf/2001.00550 in the untrainable regime. This formal analysis is consistent with the exponential cost of NQS in our numerical experiments. As a result our implementation of NQS did not outperform our high-performing tensor network algorithm.

Of course, the efficiency of NQS could be improved with sufficient future efforts. With this possibility in mind and in light of the few examples indicated by the referee, we have modified our original statement linking NSQ failure to entanglement to simply state that entanglement signals the scrambling and chaotic nature of the wave function.

Comment: As a side remark, it remains important to point out that utilizing stochastic gradient descent based solvers for neural quantum states doesn't represent state of the art, but rather higher-order solvers such as stochastic reconfiguration are utilized because only then the approach can become numerically exact with well controlled error bars.

Response: We agree that the stochastic reconfiguration (SR) is used in the NQS literature. This method multiplies the energy gradient by the inverse of the so-called geometric tensor. As a result, the method mimics the imaginary time-evolution of the wave function, projected onto the variational subspace of the NQS. We implemented such an approach and added the results to the panel FigS47d. The SR-optimized results collapse on the same curve as all other attempted approaches, which suggests that the optimization fidelity is limited by the NN expression capacity, rather than getting stuck in a local minimum. We also do the analysis of fidelities if starting from different initializations in Fig S48, suggesting no significant dependence on the initialization.

I would therefore strongly suggest to adapt these parts of the supplement accordingly.

Manuscript changes:

- In the NQS section of the supplement, following the discussion on the number of NN parameters needed to simulate $OTOC^{(2)}$ and calculation of entanglement entropies, we now state: "*This result is expected since, as Section III.C.7 shows, the wave functions produced in our $OTOC^{(2)}$ circuits have entanglement close to the maximum (with the Rényi entropy being 90% of the maximum, see Fig. S42), and are therefore highly chaotic. While, in a few cases, it is possible to find a polynomial ansatz to represent a particular volume-law entangled wave function of a ground state [citations], in the general case of a time-evolved wave function constructing such parametrization is practically hard and a typical NN architecture would require a an exponential number of parameters [citations], see also [citation]. Therefore, unlike the non-generic cases of low-entangled or low-energy wave functions, their accurate representation, in practice, requires a number of parameters of the order of* $|H|$."
- The second-to-last paragraph of the NQS section of the supplement now includes a summary of our attempt at improving NQS simulation of our circuits using stochastic gradient descent, which did not yield much different result compared with the optimizers we have already tried.
- We added a final paragraph of the NQS supplement reading: "*In conclusion, while not a formal proof, our foregoing analysis indicates that an efficient implementation of NQS algorithms outperforming TN contraction for simulating generically chaotic circuits such as $OTOC^{(2)}$ has significant practical challenges. The results do not preclude more efficient NQS optimization routines than what we have attempted here, which we leave as a subject of future work.*"

Response to editorial comments

Comment: I like your title a lot: it's balanced and captures well what you are doing. I am only wondering if we could consider a variant that suggests a bit more directly that this is paper reporting experimental observations. Maybe "Interference at the edge of the ergodic dynamics of a quantum processor"? Just a thought, feel free to tweak this or if you can't find an alternative stick to your current one.

Response: We appreciate the suggestion! Indeed the original title doesn't emphasize the experimental nature of this paper. After some discussion, we decided to try "Observation of constructive interference at the edge of quantum ergodicity". Hope this addresses the concern.

Comment: Your abstract works well but I would suggest a couple of changes for accessibility and style. The first is that right now you don't have a clear separation between the context/state of the art and the problem you're tackling and the new results you present. As a matter of style, the transition should be marked by a sentence Here we ... The second point is that the final sentence will be hard to understand for non-experts, as the link between OTOC(2) and advantage is not obvious nor explained.

Response: Thank you for the suggested abstract! We have rewritten the abstract following your flow and the suggestions above. Hope it now addresses the points you raised!

<u>Other relevant notes:</u>

- We are actively working on making the tensor contraction optimizer used to arrive at the classical simulation costs of the experiment public. A github identifier has been included in the code availability section, which should become active before publication.