Reviewers' comments:

Reviewer #1 (Remarks to the Author):

In the manuscript, Figueroa-Romero, Pollock and Modi study how an interacting quantum system can become "memoryless" despite undergoing unitary dynamics. This question is related to foundational issues in statistical mechanics and has been addressed before in different contexts, which are addressed in the references mentioned in this paper. The present work adds constructively to the discussion by introducing a definition of distance between a given channel and the closest "Markovian" channel in Sec. B. Their main result in Sec. C is to prove that this definition is satisfied for a physically well-motivated class of open system dynamics where the unitary evolution on the system and environment well-approximates a t-design. In Sec. D, they discuss examples of random unitary circuit constructions that satisfy the conditions of the theorem. The appendixes of the paper contain the proof of the main theorem in Sec. C and some other details relevant to the other sections.

Overall I found the paper to be well written and insightful and appropriate for eventual publication in this journal. My main difficulty in reviewing the paper is related to the proof of the theorem, which I have not been able to carefully check. Part of this is my own lack of expertise, but I think the authors could also improve the presentation to make the proof easier to follow. My main suggestion is to give an informal statement of the proof in the main text. In the appendix, they should provide a formal statement in addition to a sequence of formal definitions, lemmas and propositions that lay out the logical structure of the proof.

My detailed comments below primarily relate to presentation/discussion issues that I would find it helpful to have clarified:

Detailed comments:

1) On pg 1, first column, they write “This is akin to the troubling assumption of a-priori distributions in classical statistical physics.” I was not familiar with this assumption or what it means. It would be helpful to expand this discussion to at least define the terms.

2) On pg 4, second column, I found the statement of the theorem to be too technical given how the paper flowed up to that point. I would recommend giving an informal statement of the theorem, with this more formal description reserved for the appendix.

3) On pg 5, first column: “The brilliance in this construction …” The word brilliance was distracting.

4) On pg 6, first column: The sentence that begins “This shows, in principle, how simple …” is too long. It should be cut up into several smaller sentences to explain what the authors mean.

5) On pg 6, first column: “We have shown how physical quantum processes Markovianize.” This seems like too strong/simplified a summary of the results. I would suggest the wording “Markovianize in simplified toy models”.

6) On pg 6, first column. The authors discuss carbon atoms, but the results seem so far removed from such physical considerations that it is rather distracting and misleading to bring up the

connections to carbon atoms, atoms and molecules.

7) On pg 6, second column. They discuss Ref. 9, without first summarizing the result of Ref 9. It would be helpful to the general reader to first discuss what Ref. 9 is about. Similarly for Ref. 77.

8) On pg 7, first column. The sentence that begins “The definition in Eq. (10) …” is too long and convoluted. It should be broken down.

9) On pg 7, first column. The sentence “A particularly useful relation to Eq. (3) is …”. I did not understand why this relation was useful. The authors should expand the discussion to explain this point.

10) On pg 7, second column. The sentence “The function alpha_sigma is known as the …”. These concepts were unfamiliar to me. It would be helpful to define the terms and expand this discussion.

11) On pg 8, second column. The authors write “In Lemma 5.2 of [10]”. The authors should restate this lemma and define the terms.

12) On pg 10, first column. There is a typo in the sentence “Fro the fourth line …”

Reviewer #2 (Remarks to the Author):

This is a clearly-written paper that presents a rigorous result on the convergence of certain quantum processes, involving a “system” and an “environment”, to Markov processes for the “system”. It appears to be a careful piece of work (I have not checked the details of the proofs) strengthening a previous result by the same authors in Ref 9. The general topic is interesting. My reservations, however, are about the physical significance of the result and whether this is sufficiently carefully described. The significance was hard for me to fully judge for reasons mentioned below. I certainly think that some of the claims made about the significance of the result need to be more carefully stated. The manuscript also requires some additional explication of the meaning of the bound derived here.

The authors make fairly strong claims for the physical significance of their result, for example “We have shown how physical quantum processes Markovianize, i.e. forget the past.” To begin with, this generality is certainly not justified. The result applies to a very special class of systems in which each time step involves a strongly random unitary applied to the entire (system plus environment) Hilbert space. The class of random unitaries allowed is broader than in Ref 9, but still obeys extremely strong conditions that are unlikely to be satisfied in any natural experiment.

This is not in itself a criticism, as the systems studied can still be interesting in principle. However, it means that this work is very far from showing *in general* how quantum processes forget the past. In the present work,

- The system plus environment are essentially completely scrambled in between any two measurements. In an experiment, this would mean that the physical time interval between two measurements would have to grow with the size of the environment.

- The evolution is with a different, random, unitary in every time step.
- There is no spatial locality.
- The equilibrium state of the dynamics is at infinite temperature.

- If I understand correctly (I would be glad to be corrected if I have misunderstood), the processes described in this paper all converge to the *trivial* Markov process (which has Markov order 0) when the size of the environment is large.

The sentence in the conclusions beginning “Almost all processes” seems to contradict the last point above, so I would be glad to have this explained. If the point I have made above is correct, then the term “Markovianization” in the title seems to me misleading, since the results concern only a specific trivial Markov process.

There is a huge and rapidly growing literature on thermalization, which this paper does not really make contact with. The authors briefly comment that thermalization is closely related to the topic they address. But in fact thermalization is a much more general route to Markovianization than what is discussed here, in principle allowing all the restrictions above to be relaxed. For example, much recent literature makes use of random circuits, but with a spatially local structure, and (non-rigorous) results are also available for specific systems, without requiring randomness.

So, given the specific nature of the present models, it seems to me that the interest of the results depends on how strong the results obtained here are (as noted below, minimal results are trivial to obtain).

How strong the basic result is is not obvious to me. Partly this may be because I lack familiarity with the relevant preceding papers. However, I think it is also partly a defect of the presentation.

The fact that applying a random unitary to a large number of qubits thermalizes subsystems (implying that temporal correlations vanish in an evolution with multiple such unitaries) is by now well-known. Indeed it is almost trivial to use the 2-design property to show that, in the limit where the “environment” becomes large, the density matrix of the system at a given time t becomes trivial and independent of the density matrix at the previous time step.

This manuscript of course goes beyond this simple argument. For example, it uses a much more sophisticated measure of the proximity to Markovianity, carefully considers multi-time correlations, and considers t-designs for different t.

However, the end result is a complicated bound, with many terms (which involves an optimization over a parameter) and the authors do not give much intuition for the behavior of this bound. For example, the reader might ask:

- How tight is the scaling with system size, environment size, t-design order, and the parameter k?
- What is the asymptotic scaling with these quantities?
- Is the diamond norm measure really the most natural one? If I understand correctly, it assumes the experimenter is able to make arbitrarily complicated “entangled measurements” using entangled ancillas. What if, for example, only a simple sequence of projective measurements was allowed? Would this fundamentally change things?
- Are there simple physical arguments for the scaling with the various parameters mentioned above?

Are the scalings what would be expected?

If I am right in saying that the fact of Markovianization in the limit of *infinite* environment size is almost trivial, then all the interest is in how strong a bound can be given. So the above points do not seem like mere details.

I was also confused about the status of the design parameter t: in which parts of the manuscript is this assumed to be large, and why? (The last paragraph of p5 contains typos.)
Footnote 21 seems important and should be in the text body.

As a more minor point, the discussion of the toy model of particles in a box is imprecise. This is described as a physical model of particles moving in a box. Really it is a model where random unitaries are applied to qubits, and a loose *analogy* is made with collisions between particles carrying a qubit degree of freedom. This analogy requires that the positions of particles are classical rather than quantum mechanical. It also replaces their classical dynamics with trivial uncorrelated randomness, so does not really treat the particles as evolving “autonomously”: randomness is put in by hand at each time step.

Reviewer #3 (Remarks to the Author):

The article investigates the emergence of Markovian behaviour (in the sense of the process tensor) from complex quantum dynamics. As far as I understand, the principal motivation here is to extend a previous result of the same authors (Ref. [9], which shows that Haar-random evolutions are almost always Markovian) to cover more realistic evolutions, incorporating physical constraints such as locality, finite energies/timescales etc. To that end, the authors focus on a particular class of evolutions described by t-designs, for which they are able to rigorously bound the degree of non-Markovianity. A specific example of a random circuit that realises an approximate t-design is also given.

The paper considers a topic of foundational importance and is also very timely given the recent interest in information scrambling, thermalisation, and Markovianity across several different communities in physics research. The results are novel, interesting and clearly presented, and the paper is very well written. I can therefore warmly recommend publication. I do have a few minor comments on presentation for the authors to consider.

1. I think there is a typo in the definition of \Lambda below Eq. (1). Presumably the indices should run from x_0 to x_k.

2. p. 3 bottom left. The authors refer to "artificially refreshing the environment (i.e. assuming an infinite bath) *as well as* the Born-Markov approximation" (I have paraphrased slightly). But aren't those two things (refreshing & Born-Markov) basically the same assumption? For example, if you literally refresh the bath (e.g. via a "collision" model) then the Born-Markov approximation is redundant.

3. p.4 left column. The introduction and motivation of t-designs was a bit confusing for me. (I confess to a complete lack of familiarity with t-designs.) The authors refer to the cartoon of Fig. 2, pointing out that after a large number of pairwise interactions the evolution should randomise and approach

Haar-random dynamics. This makes sense. Yet the definition in Eq. (4) seems to formalise precisely the opposite statement! Namely, that a t-design looks Haar-random only for a relatively small number of repetitions s <t. It would help to clarify the physical motivation for considering a t-design here.

4. Possibly related to my previous comment/misunderstanding. How should I think about the significance of the parameter t in Fig. 5? I understand from reading the Discussion that the gate depth D is a proxy for the time duration after which the process should Markovianize. So is t somehow a proxy for the complexity of the underlying evolution? What kind of physical situation are large (or small) values of t supposed to represent? If possible, some further words of clarification might help here.

Mark T. Mitchison

Reviewer #4 (Remarks to the Author):

The manuscript by Figueroa-Romero, Pollock and Modi shows that if a process is not interrupted with too many interventions it will look markovian (provided some reasonable assumptions on the dimension of the environment hold). The authors show that is the process is described by an approximate unitary design one can bound the amount (size) of non-markovian memory! That's really cool.

The general area explored in this manuscript is of great interest to people in quantum information and it is possible these results could find applications in condensed matter, quantum chaos, and the toy models of quantum gravity explored in quantum information. This is carefully executed work and really interesting. I'm very positive about the manuscript and I recommend it should be accepted with the opportunity for the authors to make minor revisions should they desire.

Below I make some small suggestions and raise some small issues that the authors may wish to address. My recommendation of acceptance is not conditional on the authors addressing the issues below. The authors are the experts in this field, and they should decide if the points I raise below have merit.

- In figure 2 is there a better picture for the toy model? A naïve translation of that figure would be a collection of N particles in three-dimensional space with appropriate position and momentum operators. Maybe the author's model is closer to a collection of coins with pairwise interactions? Again if this is not helpful please ignore it.
- Would it make more sense to log the y axis as well on figure 3? Maybe it doesn't look great?
- Perhaps Section II A should be renamed to "Discrete-time quantum stochastic processes" as continuous-time processes are not studied. Moreover it is well known that there are discrete-time processes that *do not* have a sensible continuous-time limit and vice versa.
- I was curious if the authors had considered the extension of these results to continuous variables systems. Often, in quantum optics, authors have argued that taking the limit of dimension -> \infty is sensible. The mathematicians and mathematical physicists usually point to the C* algebraic results as the correct formalization. I believe there has been a little work in the C* algebra community on quantum markov processes (by the usual suspects e.g. L. Accardi, A. Holevo, V. P. Belavkin, F. Fagnola, D. Petz, J Gough, B. Kümmerer, R. Streater, I. Wilde, K. R. Parthasarathy etc), likely of a far

less general nature that the investigations of the authors. As an interested reader I was curious to know how applicable the current results are in general do they hold for bounded operators on L2R and also unbounded? Perhaps it might be worth mentioning in the conclusion?

- Is it clear that the Born approximation usually invoked is inequivalent to requiring a high dimensional environment? Or that requiring weak coupling is inequivalent to not too many interventions? From the requirement that d_E>> d_S^{2*# intervetions+1} and equation 17, it seems hard to conclusively argue they are different. This narrative is stressed throughout the abstract, main text, and in the appendix e.g. “The significance of Eq. (15) is thus that quantum processes with not too many interventions in high dimensional environments will look to be almost Markovian with high probability. “
- I’m not sure what the history of this field is, perhaps there are some heated debates? I’m mentioning this because the paper comes across as very polemical e.g. the abstract, introduction (specifically paragraphs 2 and 3), section II A (paragraph 2). As someone who doesn’t have a vested interest in the area, it’s quite off putting. Some of the statements seem problematic to me. For example consider the quote “The derivations of such processes are fraught with ad-hoc assumptions such as artificially refreshing the environment between time-steps (i.e., assumption of an infinite bath), as well as adhering to approximations such as Born-Markov.” Physical systems that (i) obey these kinds of descriptions or (ii) are intentionally engineered to obey these descriptions have been routinely built in labs and measured daily for close to 3 decades. In contrast, the specific circuits authors suggest to perform markovianzation are only just starting to be built. I’m certain that the authors have a clear idea of what they mean and could expand the text to easily rebut my above comments. But that is missing my point. I am reading your paper: tell me what is awesome about your work! I'm less interested in what is substandard in other's work. You have to ask yourself does the polemical tone contribute to the message of the paper?

Response to Reviewer #1 – Reviewer's comments quoted in italics

In the manuscript, Figueroa-Romero, Pollock and Modi study how an interacting quantum system can become "memoryless" despite undergoing unitary dynamics. This question is related to foundational issues in statistical mechanics and has been addressed before in different contexts, which are addressed in the references mentioned in this paper. The present work adds constructively to the discussion by introducing a definition of distance between a given channel and the closest "Markovian" channel in Sec. B. Their main result in Sec. C is to prove that this definition is satisfied for a physically well-motivated class of open system dynamics where the unitary evolution on the system and environment well-approximates a t-design. In Sec. D, they discuss examples of random unitary circuit constructions that satisfy the conditions of the theorem. The appendixes of the paper contain the proof of the main theorem in Sec. C and some other details relevant to the other sections. Overall I found the paper to be well written and insightful and appropriate for eventual publication in this journal.

We thank the Reviewer for an accurate reading of our work and for their support for our manuscript.

My main difficulty in reviewing the paper is related to the proof of the theorem, which I have not been able to carefully check. Part of this is my own lack of expertise, but I think the authors could also improve the presentation to make the proof easier to follow. My main suggestion is to give an informal statement of the proof in the main text. In the appendix, they should provide a formal statement in addition to a sequence of formal definitions, lemmas and propositions that lay out the logical structure of the proof.

We have now addressed this and rearranged the proof of the main theorem so that it's easier to follow.

My detailed comments below primarily relate to presentation/discussion issues that I would find it helpful to have clarified. Detailed comments:

1) On pg 1, first column, they write "This is akin to the troubling assumption of a-priori distributions in classical statistical physics." I was not familiar with this assumption or what it means. It would be helpful to expand this discussion to at least define the terms.

We have now added that this refers to the so-called "Fundamental Postulate of Statistical Mechanics", also known as the "Principle of Equal A Priori Probabilities".

2) On pg 4, second column, I found the statement of the theorem to be too technical given how the paper flowed up to that point. I would recommend giving an informal statement of the theorem, with this more formal description reserved for the appendix.

We thank the reviewer for their comment; we have decided to keep a formal statement, with modifications, as detailed in this response, to the discussion around it which should make it clearer.

3) On pg 5, first column: "The brilliance in this construction ..." The word brilliance was distracting.

We agree and we have rephrased this.

4) On pg 6, first column: The sentence that begins "This shows, in principle, how simple ..." is too long. It should be cut up into several smaller sentences to explain what the authors mean.

We have now broken up this sentence.

5) On pg 6, first column: "We have shown how physical quantum processes Markovianize." This seems like too strong/simplified a summary of the results. I would suggest the wording "Markovianize in simplified toy models".

We agree and we have rephrased this.

6) On pg 6, first column. The authors discuss carbon atoms, but the results seem so far removed from such physical considerations that it is rather distracting and misleading to bring up the connections to carbon atoms, atoms and molecules.

This line is just meant to evoke the foundational importance of the idea of forgetfulness addressed in our work. It is also meant to help the reader develop intuition for the fundamental physics that we touch on in the paper.

7) On pg 6, second column. They discuss Ref. 9, without first summarizing the result of Ref 9. It would be helpful to the general reader to first discuss what Ref. 9 is about. Similarly for Ref. 77.

Ref. 9 is referred to throughout the article, starting on page 1 second column. We now point out what the result in Ref. 9 says as well on pg 6.

8) On pg 7, first column. The sentence that begins "The definition in Eq. (10) ..." is too long and convoluted. It should be broken down.

We agree, and have now split this sentence.

9) On pg 7, first column. The sentence "A particularly useful relation to Eq. (3) is ...". I did not understand why this relation was useful. The authors should expand the discussion to explain this point.

We now explicitly mention that it's useful in the sense of allowing to directly bound the diamond norm once any Schatten norm is known.

10) On pg 7, second column. The sentence "The function α_σ is known as the ...". These concepts were unfamiliar to me. It would be helpful to define the terms and expand this discussion.

We have added a sentence for clarification and expanded the discussion within an Appendix subsection.

11) On pg 8, second column. The authors write "In Lemma 5.2 of [10]". The authors should restate this lemma and define the terms.

As the Reviewer points out, we have now included this for completeness.

12) On pg 10, first column. There is a typo in the sentence "For the fourth line ..."

We have fixed this and we thank again the Reviewer for their very detailed comments.

Response to Reviewer #2 – Reviewer's comments quoted in italics

This is a clearly-written paper that presents a rigorous result on the convergence of certain quantum processes, involving a "system" and an "environment", to Markov processes for the "system". It appears to be a careful piece of work (I have not checked the details of the proofs) strengthening a previous result by the same authors in Ref 9. The general topic is interesting. My reservations, however, are about the physical significance of the result and whether this is sufficiently carefully described. The significance was hard for me to fully judge for reasons mentioned below. I certainly think that some of the claims made about the significance of the result need to be more carefully stated. The manuscript also requires some additional explication of the meaning of the bound derived here.

We thank the Reviewer for their comments; we address their concerns point by point below.

*The authors make fairly strong claims for the physical significance of their result, for example "We have shown how physical quantum processes Markovianize, i.e. forget the past." To begin with, this generality is certainly not justified. The result applies to a very special class of systems in which each time step involves a strongly random unitary applied to the entire (system plus environment) Hilbert space. The class of random unitaries allowed is broader than in Ref 9, but still obeys extremely strong conditions that are unlikely to be satisfied in any natural experiment. This is not in itself a criticism, as the systems studied can still be interesting in principle. However, it means that this work is very far from showing *in general* how quantum processes forget the past.*

We agree with the Reviewer; we have now stressed the significance of our work while making clear the class of dynamics that we specifically addressed. However, as we exemplify in Section D, many evolutions, that can be broken up into two types of alternating interactions between subsystems, will fit the general requirements of our theorem. This is why we claim that a large subclass of Hamiltonians will fall under our class. We finally point out that unitary t designs are probability distributions over a finite set of unitaries such that t copies of these are indistinguishable from those randomly sampled from the uniform (Haar) measure. That is, the class of evolutions we consider are sampled from this ensemble, but need not be random themselves. That is, if a single process from this ensemble is run many times, Markovianization will still occur with high probability.

In the present work,
- The system plus environment are essentially completely scrambled in between any two measurements. In an experiment, this would mean that the physical time interval between two measurements would have to grow with the size of the environment.

The Reviewer is correct in that we consider a class of systems that rapidly scramble information, namely, unitary designs; however, with this approach we are precisely stepping away from *fully* random (i.e. Haar random) dynamics, implying that this scrambling occurs in a weaker way.

- The evolution is with a different, random, unitary in every time step.

We stick to the case of different unitaries at every time-step for concreteness; we stress again that these need not be random, as we only require these to belong to an approximate unitary design, which by definition only replicates a finite amount of statistical moments of the Haar measure.

- There is no spatial locality.

Indeed, we don't place an a-priori restriction on spatial locality.

- The equilibrium state of the dynamics is at infinite temperature.

Our equivalent of an equilibrium state is *some* Markovian process $\Upsilon^{(\mathrm{M})}$ (these are always of the form given in Eq. (2)) *that is closest* to the process Υ being considered, as it enters in Eq. (3). Notice that this implies that not all such $\Upsilon^{(\mathrm{M})}$ will necessarily be the same; for example, if instead of the black diamond norm we had chosen the relative entropy between both processes, we could have easily seen that the minimizer of such distance is always a product of marginals of Υ, i.e. $\Upsilon^{(\mathrm{M})}_{\mathrm{min}} = \Upsilon_{\mathsf{S}} \otimes \Upsilon_{\mathrm{ancilla}_1} \otimes \Upsilon_{\mathrm{ancilla}_2} \otimes \cdots$. We only employ the maximally mixed state (here a maximally noisy process), which is a particular case of the definition in Eq.(2), to put a bound on the general non-Markovianity defined by Eq. (3).

*- If I understand correctly (I would be glad to be corrected if I have misunderstood), the processes described in this paper all converge to the *trivial* Markov process (which has Markov order 0) when the size of the environment is large.*

This is essentially correct but it can be phrased more accurately as: we consider a class of quantum processes (those generated by approximate unitary designs) and show that these will look almost Markovian with high probability whenever the (finite) environment is large, the amount of time-steps is small, and the design is large and accurate (i.e. large design degree t and small error ϵ). For each unitary process the the Markov process will be unique but will be close to the trivial one.

The sentence in the conclusions beginning "Almost all processes" seems to contradict the last point above, so I would be glad to have this explained. If the point I have made above is correct, then the term "Markovianization" in the title seems to me misleading, since the results concern only a specific trivial Markov process.

With the phrase "Almost all", we mean with *high probability* a process that is sampled from t-design will be "almost Markovian". With "almost Markovian" we mean *close* to Markov order zero as per the distance defined in Eq. (3). As we explain a couple of paragraphs above, when a process Markovianizes, it will do so in general to a different Markovian process. This Markovianization to some Markovian process that minimizes Eq. (3) is already implied whenever the process is already close to the trivial one.

There is a huge and rapidly growing literature on thermalization, which this paper does not really make contact with. The authors briefly comment that thermalization is closely related to the topic they address. But in fact thermalization is a much more general route to Markovianization than what is discussed here, in principle allowing all the restrictions above to be relaxed. For example, much recent literature makes use of random circuits, but with a spatially local structure, and (non-rigorous) results are also available for specific systems, without requiring randomness.

The Reviewer is correct about the vastness of results currently available. As we point out, however, the relation between Markovianization and equilibration/thermalization is not entirely clear yet. We also consider one possible approach to Markovianization with physically motivated systems using unitary designs and no further restrictions, but there is indeed ample room to consider more specific systems. Moreover, as we point out in the paper, Markovianization leads to unique steady states, which can be a mechanism for thermalisation. Furthermore, a system may look thermal and still possess higher-order correlations in time. Thus, we believe that Markovianization must occur before thermalisation and is a stronger notion.

So, given the specific nature of the present models, it seems to me that the interest of the results depends on how strong the results obtained here are (as noted below, minimal results are trivial to obtain). How strong the basic result is is not obvious to me. Partly this may be because I lack familiarity with the relevant preceding papers. However, I think it is also partly a defect of the presentation.

We have now highlighted the reach of our results as well as the limitations of arguments based on typicality. We point out as well that Fig. 3 captures to a large extent the behaviour of the bound in our main theorem.

The fact that applying a random unitary to a large number of qubits thermalizes subsystems (implying that temporal correlations vanish in an evolution with multiple such unitaries) is by now well-known. Indeed it is almost trivial to use the 2-design property to show that, in the limit where the "environment" becomes large, the density matrix of the system at a given time t becomes trivial and independent of the density matrix at the previous time step. This manuscript of course goes beyond this simple argument. For example, it uses a much more sophisticated measure of the proximity to Markovianity, carefully considers multi-time correlations, and considers t-designs for different t. However, the end result is a complicated bound, with many terms (which involves an optimization over a parameter) and the authors do not give much intuition for the behavior of this bound.

As the Reviewer points out, the case we consider goes beyond typicality of quantum states; on the other hand, we have tried to present sufficient intuition about the bound we derived in our main theorem with the discussion following it, particularly through Fig. 3, and with the example in Section D employing a simple 2-qubit interaction circuit. We nevertheless acknowledge the Reviewer's concern and address these specific comments below.

For example, the reader might ask:

- How tight is the scaling with system size, environment size, t-design order, and the parameter k? What is the asymptotic scaling with these quantities?

We have now added a paragraph below our main theorem about these scalings.

- Is the diamond norm measure really the most natural one? If I understand correctly, it assumes the experimenter is able to make arbitrarily complicated "entangled measurements" using entangled ancillas. What if, for example, only a simple sequence of projective measurements was allowed? Would this fundamentally change things?

We now mention that when we say this is a natural choice we mean it has a clear operational definition as well being one of the most general measures, containing particular cases such as the one mentioned by the Reviewer including only projective measurements.

- Are there simple physical arguments for the scaling with the various parameters mentioned above? Are the scalings what would be expected?

We discuss this throughout the manuscript, but now we have stressed this within the paragraph on scalings mentioned above.

*If I am right in saying that the fact of Markovianization in the limit of *infinite* environment size is almost trivial, then all the interest is in how strong a bound can be given. So the above points do not seem like mere details. I was also confused about the status of the design parameter t: in which parts of the manuscript is this assumed to be large, and why? (The last paragraph of p5 contains typos.)*

First, note that an infinite environment does not imply Markovianization. The spin-boson model has such an environment and also has complex non-Markovian structures. There are no assumptions on the design parameter t; as mentioned around Eq. (4), these give a finite quantification to the statistical moments with the Haar measure and in this sense quantify complexity (as implementing a Haar random unitary requires an exponential number of two-qubit gates and random bits), thus being normally expected to give large deviation bounds for large t. We now also comment on this for further clarity.

Footnote 21 seems important and should be in the text body.

We agree and have now changed this.

*As a more minor point, the discussion of the toy model of particles in a box is imprecise. This is described as a physical model of particles moving in a box. Really it is a model where random unitaries are applied to qubits, and a loose *analogy* is made with collisions between particles carrying a qubit degree of freedom. This analogy requires that the positions of particles are classical rather than quantum mechanical. It also replaces their classical dynamics with trivial uncorrelated randomness, so does not really treat the particles as evolving "autonomously": randomness is put in by hand at each time step.*

The Reviewer is correct that we mention this model only as an analogy and in order to motivate the question of Markovianization on physically plausible systems; it is, for example, still an open question if there exist two-body local time-independent Hamiltonians generating unitary designs.

Response to Reviewer #3 – Reviewer's comments quoted in italics

The article investigates the emergence of Markovian behaviour (in the sense of the process tensor) from complex quantum dynamics. As far as I understand, the principal motivation here is to extend a previous result of the same authors (Ref. [9], which shows that Haar-random evolutions are almost always Markovian) to cover more realistic evolutions, incorporating physical constraints such as locality, finite energies/timescales etc. To that end, the authors focus on a particular class of evolutions described by t-designs, for which they are able to rigorously bound the degree of non-Markovianity. A specific example of a random circuit that realises an approximate t-design is also given.

We thank the Reviewer for an accurate assessment of our manuscript.

The paper considers a topic of foundational importance and is also very timely given the recent interest in information scrambling, thermalisation, and Markovianity across several different communities in physics research. The results are novel, interesting and clearly presented, and the paper is very well written. I can therefore warmly recommend publication. I do have a few minor comments on presentation for the authors to consider.

We also thank the Reviewer for their recommendation and support of our work.

1. I think there is a typo in the definition of Λ below Eq. (1). Presumably the indices should run from x_0 to x_k.

This is correct, we have now fixed this typo.

*2. p. 3 bottom left. The authors refer to "artificially refreshing the environment (i.e. assuming an infinite bath) *as well as* the Born-Markov approximation" (I have paraphrased slightly). But aren't those two things (refreshing & Born-Markov) basically the same assumption? For example, if you literally refresh the bath (e.g. via a "collision" model) then the Born-Markov approximation is redundant.*

This is correct and thank the Reviewer for pointing it out, we now phrase this correctly meaning that these imply each other.

3. p.4 left column. The introduction and motivation of t-designs was a bit confusing for me. (I confess to a complete lack of familiarity with t-designs.) The authors refer to the cartoon of Fig. 2, pointing out that after a large number of pairwise interactions the evolution should randomise and approach Haar-random dynamics. This makes sense. Yet the definition in Eq. (4) seems to formalise precisely the opposite statement! Namely, that a t-design looks Haar-random only for a relatively small number of repetitions $s < t$. It would help to clarify the physical motivation for considering a t-design here.

Unitary designs are ensembles of unitaries with a finite amount of statistical moments matching those with the Haar measure; in this sense, the more moments the design matches with the Haar measure, the more Haar-random dynamics the corresponding unitaries will generate. This is precisely captured quantitatively in Eq. (4) by saying that an ϵ-approximate unitary t-design will differ in all its moments up to the t^{th} one from the Haar measure by a small ϵ. The conundrum that nature is far from random but nevertheless systems in it seem to typically forget their past, is what gave us the physical motivation to employ this finite approximation to Haar randomness through unitary designs.

4. Possibly related to my previous comment/misunderstanding. How should I think about the significance of the parameter t in Fig. 5? I understand from reading the Discussion that the gate depth $\mathfrak{D}$ is a proxy for the time duration after which the process should Markovianize. So is t somehow a proxy for the complexity of the underlying evolution? What kind of physical situation are large (or small) values of t supposed to represent? If possible, some further words of clarification might help here.

This is correct; in Fig. 5, specifically, Markovianization occurs for the values of $\mathfrak{D}$ plotted in the sense of $\mathbb{P}_{\mu_{t_\varepsilon}}[\mathcal{N} \geq 0.1] \leq 0.01$ being satisfied. It is also correct that the parameter t can be directly related to the complexity of the dynamics in the sense of a minimum number of elementary gates necessary to generate the corresponding unitary operators; physically this relates to concepts such as dissipation, scrambling or chaos in the dynamics, with a higher value of t representing a more random or chaotic dynamics.

Response to Reviewer #4 – Reviewer's comments quoted in italics

The manuscript by Figueroa-Romero, Pollock and Modi shows that if a process is not interrupted with too many interventions it will look Markovian (provided some reasonable assumptions on the dimension of the environment hold). The authors show that is the process is described by an approximate unitary design, one can bound the amount (size) of non-Markovian memory! That's really cool.

We thank the Reviewer for their accurate reading of our manuscript and their supporting comments.

The general area explored in this manuscript is of great interest to people in quantum information and it is possible these results could find applications in condensed matter, quantum chaos, and the toy models of quantum gravity explored in quantum information. This is carefully executed work and really interesting. I'm very positive about the manuscript and I recommend it should be accepted with the opportunity for the authors to make minor revisions should they desire. Below I make some small suggestions and raise some small issues that the authors may wish to address. My recommendation of acceptance is not conditional on the authors addressing the issues below. The authors are the experts in this field, and they should decide if the points I raise below have merit.

We thank again the Reviewer for their support and recommendation for publication of our manuscript, we now address in detail their comments.

- In figure 2, is there a better picture for the toy model? A naïve translation of that figure would be a collection of N particles in three-dimensional space with appropriate position and momentum operators. Maybe the author's model is closer to a collection of coins with pairwise interactions? Again if this is not helpful please ignore it.

We agree that the depiction does not exactly correspond to the design circuit model; we now explicitly mention that we do not exactly study this model but that this is a cartoon evoking a physical situation resembling the two-qubit model that we describe in Section D.

Would it make more sense to log the y axis as well on figure 3? Maybe it doesn't look great?

We tried different visualizations for the plot in Fig.3 but agreed on the current one due to it displaying the polynomial decaying behaviour progressively becoming exponential in increasing design parameter t.

*Perhaps Section II A should be renamed to "Discrete-time quantum stochastic processes" as continuous-time processes are not studied. Moreover it is well known that there are discrete-time processes that *do not* have a sensible continuous-time limit and vice versa.*

We now explicitly point out that the process tensor framework for quantum stochastic processes that we employ has been shown to allow for continuous, or active, interventions. More precisely, it has been shown to satisfy a generalized extension theorem, akin to Kolmogorov's theorem for classical stochastic processes, where the minimal conditions for an underlying continuous process to exist are set, and wherein the finite sequence of interventions would become a continuous driving or control of the system.

I was curious if the authors had considered the extension of these results to continuous variables systems. Often, in quantum optics, authors have argued that taking the limit of dimension $\to \infty$ is sensible. The mathematicians and mathematical physicists usually point to the C^* algebraic results as the correct formalization. I believe there has been a little work in the C^* algebra community on quantum markov processes (by the usual suspects e.g. L. Accardi, A. Holevo, V. P. Belavkin, F. Fagnola, D. Petz, J Gough, B. Kümmerer, R. Streater, I. Wilde, K. R. Parthasarathy etc), likely of a far less general nature that the investigations of the authors. As an interested reader I was curious to know how applicable the current results are in general do they hold for bounded operators on L2R and also unbounded? Perhaps it might be worth mentioning in the conclusion?

We thank the Reviewer for pointing this out. In fact, the works due to some of the authors listed above do construct very similar structure for non-Markovian quantum processes. However, there are subtle differences. Most notably, the process tensor maps quantum processes to quantum states, which allows us to explore the statistical features of quantum processes with ease. It is almost certainly possible to do the same in the C^* algebra setting, but we leave these questions for a separate work in itself.

Is it clear that the Born approximation usually invoked is inequivalent to requiring a high dimensional environment? Or that requiring weak coupling is inequivalent to not too many interventions? From the requirement that $d_E \gg d_S^{2k+1}$ and equation 17, it seems hard to conclusively argue they are different. This narrative is stressed throughout the abstract, main text, and in the appendix e.g. "The significance of Eq. (15) is thus that quantum processes with not too many interventions in high dimensional environments will look to be almost Markovian with high probability."

Our work aims to explain the existence and emergence of Markovianity in systems that inherently (by definition) are non-Markovian. We are not arguing that the assumptions for the Born-Markov approximation are wrong or in conflict with out results, but rather we formalize and generalize these conditions, which arise naturally when processes are typical with respect to the Haar measure, or specifically as we show in this manuscript, with respect to approximate unitary designs. As we conclude the paragraph the Reviewer cites, our results are a step towards explaining "the emergence of Markovian processes without ad-hoc assumptions. . . ". The question we ask in our manuscript is how do Markovian processes emerge, i.e. how can they be assumed, on processes which are inherently non-Markovian? The Reviewer can also appreciate this by considering the case of the emergence of statistical mechanics from quantum mechanics, namely, how can microcanonical ensembles emerge without equal a-priori probabilities and purely from quantum mechanics? Ours is the analogue question for quantum processes.

I'm not sure what the history of this field is, perhaps there are some heated debates? I'm mentioning this because the paper comes across as very polemical e.g. the abstract, introduction *(specifically paragraphs 2 and 3), section II A (paragraph 2). As someone who doesn't have a vested interest in the area, it's quite off putting. Some of the statements seem problematic to me. For example consider the quote "The derivations of such processes are fraught with ad-hoc assumptions such as artificially refreshing the environment between time-steps (i.e., assumption of an infinite bath), as well as adhering to approximations such as Born-Markov." Physical systems that (i) obey these kinds of descriptions or (ii) are intentionally engineered to obey these descriptions have been routinely built in labs and measured daily for close to 3 decades. In contrast, the specific circuits authors suggest to perform markovianization are only just starting to be built. I'm certain that the authors have a clear idea of what they mean and could expand the text to easily rebut my above comments. But that is missing my point. I am reading your paper: tell me what is awesome about your work! I'm less interested in what is substandard in other's work. You have to ask yourself does the polemical tone contribute to the message of the paper?*

We thank the Reviewer for pointing this out. It was certainly not our intention to come across as antagonistic, but rather to motivate that these are assumptions and approximations which raise foundational (as well as practical) questions that should be addressed. In fact, tools like the quantum master equations are used by a very wide array of communities, far beyond physics. Many of the researchers are fully aware of the ad hoc nature of the assumptions, but continue to use the tools in regime beyond where the assumptions will hold; this is plainly because they work. However, there must be physical justifications for why these assumptions work and a regime where they don't. One of our work's motivation is to close this gap. We highlight that it is precisely the wide applicability and success of the Born-Markov approximation which raises the question of the emergence and typicality of Markovianity. We have now rephrased the sentence cited by the Reviewer.

Reviewers' comments:

Reviewer #1 (Remarks to the Author):

The authors have satisfactorily addressed all my comments and, in my view, significantly improved the manuscript. I now recommend the paper for publication.

Reviewer #2 (Remarks to the Author):

The authors have made some minor changes that slightly clarify their bound, which is a minor improvement. I still have two queries, one about presentation and one about the motivation for the calculation, labeled (1) and (2) below.

The authors talk throughout about "Markovianization". If I understand them correctly, this means proximity to a process of Markov order 1. Indeed in the paper they say "Here, m is the Markov order; when m = 1 the process is called Markovian".

Therefore I assume that in the following comment in their response letter, "zero" is a typo for "one": "With “almost Markovian” we mean close to Markov order zero as per the distance defined in Eq. (3)."

(1) The models the authors study converge to the trivial process, of Markov order 0, in the limit. Given that the whole discussion is framed by the authors in terms of how close you are to an arbitrary Markov process (a process of Markov order 1), this still seems to me likely to confuse the reader. It should be clarified in the text.

In other words: since the basic mechanism is convergence to the trivial process, it would be helpful to state that directly. Otherwise, the authors run the risk of making the results seem more mysterious than they are.

(2) The norm that the authors use is based on the distance to the closest Markov process. It involves an optimization over all Markov processes.

Since the dynamics here converge to a specific process (the trivial process with Markov order 0) I do not see the utility of this more complex measure.

Why not just study the distance to the reference (trivial) process?

- The authors state in the reply:

"we stress again that these need not be random, as we only require these to belong to an approximate unitary design, which by definition only replicates a finite amount of statistical moments of the Haar measure."

I assume this is a typo for "need not be Haar random". A unitary design a probability distribution over unitaries, i.e. the unitaries the authors study are random by definition.

- In the first review I stated "If I am right in saying that the fact of Markovianization in the limit of *infinite* environment size is almost trivial, then all the interest is in how strong a bound can be given." To clarify, this referred to the limit d_E ->infinity in a model where the dynamics is given by a 2-design, i.e. a model like those discussed in this paper. I was not referring to arbitrary models such as the spin-boson model mentioned by the authors in their response.

Reviewer #4 (Remarks to the Author):

I am happy with all of the changes the authors have made.

Response to Reviewer #2 – Reviewer's comments quoted in italics

The authors have made some minor changes that slightly clarify their bound, which is a minor improvement. I still have two queries, one about presentation and one about the motivation for the calculation, labeled (1) and (2) below. The authors talk throughout about "Markovianization". If I understand them correctly, this means proximity to a process of Markov order 1. Indeed in the paper they say "Here, m is the Markov order; when m = 1 the process is called Markovian". Therefore I assume that in the following comment in their response letter, "zero" is a typo for "one": "With "almost Markovian" we mean close to Markov order zero as per the distance defined in Eq. (3)."

We thank the Reviewer again for their comments. What they point out is correct, it should have said "order one" in our previous response.

(1) The models the authors study converge to the trivial process, of Markov order 0, in the limit. Given that the whole discussion is framed by the authors in terms of how close you are to an arbitrary Markov process (a process of Markov order 1), this still seems to me likely to confuse the reader. It should be clarified in the text. In other words: since the basic mechanism is convergence to the trivial process, it would be helpful to state that directly. Otherwise, the authors run the risk of making the results seem more mysterious than they are.

We would respectfully disagree on this; the main message of our result is that with high probability almost all processes undergone in high dimensional environments with complex enough dynamics are *almost* Markovian. The Reviewer is right about the convergence to Markov order zero but this rather a consequence in the mentioned infinite dimensional limit, and at most it is only the means that we used in turn implying convergence to Markov order one processes.

(2) The norm that the authors use is based on the distance to the closest Markov process. It involves an optimization over all Markov processes. Since the dynamics here converge to a specific process (the trivial process with Markov order 0) I do not see the utility of this more complex measure. Why not just study the distance to the reference (trivial) process?

As we mention above, we are interested in the distance with the trivial process insofar that it can bound the distance with non-trivial, Markov order one, processes. The main reason for this is that trivial processes are just that, noise, while Markov order one encompasses all other ubiquitous phenomena in science commonly labelled as Markovian. In a sense it is true that we are overshooting when we get convergence to the trivial process but the importance of this is that it implies convergence of the operationally motivated measure of non-Markovianity in Eq. (3) that we employ.

- The authors state in the reply: "we stress again that these need not be random, as we only require these to belong to an approximate unitary design, which by definition only replicates a finite amount of statistical moments of the Haar measure." I assume this is a typo for "need not be Haar random". A unitary design a probability distribution over unitaries, i.e. the unitaries the authors study are random by definition.

This is correct, we meant uniformly random or Haar random.

*- In the first review I stated "If I am right in saying that the fact of Markovianization in the limit of *infinite* environment size is almost trivial, then all the interest is in how strong a bound can be given." To clarify, this referred to the limit $d_E \to \infty$ in a model where the dynamics is given by a 2-design, i.e. a model like those discussed in this paper. I was not referring to arbitrary models such as the spin-boson model mentioned by the authors in their response.*

We thank the Reviewer for the clarification, then in that case we agree our comment about the spin-boson was redundant.