nature portfolio

Peer Review File

On the potential and limitations of quantum extreme learning machines

Open Access This file is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. In the cases where the authors are anonymous, such as is the case for the reports of anonymous peer reviewers, author attribution should be to 'Anonymous Referee' followed by a clear attribution to the source work. The images or other third party material in this file are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

Reviewers' comments:

Reviewer #1 (Remarks to the Author):

This manuscript addresses a supervised quantum machine learning approach known as quantum extreme learning machine (QELM) aiming to establish its limitations, e.g. for the sampling noise in measurements, and its potential, e.g. the possibility to use multiple injections to implement non-linear functions of the input state. This is a timely subject and the manuscript presents insightful results. There are 3 main sections covering formal analysis, and numerical results about the performance of this method for different tasks in presence of either single or multiple (repeated) injections of the input state.
The major contribution is that the authors describe QELM in the language of channels and POVM, providing a useful framework for the community working in this field. As for the numerical results some clarifications are needed, but overall the manuscript represents an interesting contribution and I am willing to recommend it for publication after the authors improve their presentation addressing the issues detailed below.

MAIN ISSUES

1) While the theory section is well detailed, the discussion of the results in Sections III an IV is rather limited, not covering/analyzing all the presented results. For instance Fig.4 and Fig.6(b) are not even discussed.
Some specific recommendations to improve these sections are to
- clarify the meaning of an error of order 1000 (!) in estimating sigma_x (of order 1) in Fig.3(d). Maybe related to this (at least) singular result, the authors could also comment the effect of changing the random unitary (in this and the other figures) .
- in Fig.3 explicitly state if this is the case of the scenario 1 and for which qudit size;
- display the plot of the ideal case in Fig.3 to see the deviation from it;
- clarify if/that the x axis In Fig.3 refers to the output size; the names “number of samples” and “number of measurements” can be misleading. Does the latter correspond to m in sect II?
-in Fig.3 it should be clarified that all results are obtained for the test data (is it the case?); Are the training and test samples set to the size of the number of samples as indicated in the caption or is it just a problem of notation? Anyway provide the numbers.
- comment on the dependence of these results when changing the Hamiltonian in Fig.3;
- clarify what does exactly represent each dot in Fig.4;
- the authors actually do not draw any conclusion on Fig.4. Which is the conveyed message?
- Overall the results in Figs 3 and 4 are really hard to see and could be represented in larger sizes.
- Fig.5 does not really add too much insight (and is rather large);
- clarify the meaning of the peak in 9 in Fig 6 (I could not understand it from the comment at the end of the section);
- add a comment about the importance/relevance of Fig.6(b); comment on the steps arising for the exponential task as well as the “flatness” of the other one.

2) The authors introduce a technical discussion about state pseudo-inverse and condition number in Section IID and Fig.2 but the main novel conclusions and implications of this analysis for QELM are not really clear, also looking at the growth with N and peak in Fig.3(a). Which is the use of this (peaked) figure of merit in order to assess the QELM performance? Given the unclear relevance of these findings, this long discussion on condition number could probably be shortened.
As a comment, building on this discussion, do the authors consider any regularization procedure, as commonly considered in classical settings?

3) QRC is often mentioned but actually neither the theory framework, nor the results, cover this case that seems actually rather different. The recommendation is to remove unrelated sentences in abstract and introduction referring to results on QRC (anticipating results about “ELM or QRC” while

only ELM is covered). Also the captions of table I and Fig.5 make a reference to QRC but do not address this case. Considering that QRC significantly departs from QELM the fact that it “can also be analysed along similar lines” remains to be proved. Finally, the statement about the memory in the conclusions is hard to follow looking at the contents of this work...

MINOR ISSUES

Missing axis in Fig.2.

In the introduction when referring to [24] maybe the authors meant “implementation of nonlinear input-output maps with memory” on a NISQ device.

The authors write that “characterisation of the class of tasks that are accomplishable through QRC- or QELM [...] is lacking”, but this is actually addressed in most of the references they cite considering a broad spectrum of tasks. Also in the abstract they write about a lacking characterization of potential and limitations but actually these aspects have been addressed from different perspectives in several works, starting from the first one [10].

Notation:
In Section C the observable O is defined from the target, while after eq.(5) the observable correspond to the QELM output. Should the notation be disambiguated?
Also the use of vectors instead of single components in Sections B and C is not always clear.

In Section III the authors introduce 3 scenarios for QELMs. Some references to their use in the literature should be included. Furthermore the following examples could make explicit reference to corresponding relative scenario (1 or 2 or 3) to guide the reader, e.g. in Fig.3 or 4.

Reviewer #2 (Remarks to the Author):

The manuscript shows that reservoir computer and extreme learning machines can be used for quantum feature reconstructions. This is an interesting and timely work. However, I think there are some required improvements before it can be considered for publication.

First, the claims in the introduction part of manuscript are made very general and is not in agreement with the results shown in the rest of the manuscript. This must be addressed and need to be more accurate.

For instance, see this paragraph: “In this paper, we make significant progress in this direction. First, we show that the problem of reconstructing features of a quantum state via an ELM-like setup can be viewed as a linear regression task on the measurement probabilities produced by a suitable positive operator valued measurement (POVM). The key observation is that the probability distribution corresponding to an arbitrary measurement of a quantum state is linear in the input density matrix….”

In this paragraph, the claim is that the problem of reconstructing features of a quantum state via an ELM-like setup can be viewed as a linear regression task. This has been shown in several previous works. These previous works must be properly acknowledged. In fact, quantum reservoir networks have been successfully used for quantum state tomography, which is arguably the most complete for of feature reconstruction of quantum state. Given that these published works are directly related to the claims of the paper, I would suggest the authors to not claim these as their own original finding and cite the relevant papers in this paragraph. For instance, see these Refs [Phys. Rev. Lett. 127, 260401 (2021), and IEEE TNNLS, 32, 3148-3155 (2021)].

However, I would like to say that the study reported in the subsequent paragraph is important and original contribution. This paragraph can be more highlighted.

In Figure 2, the axes are not labelled.

Figure 6 is missing in the manuscript.

Reviewer #3 (Remarks to the Author):

In the manuscript "On the potential and limitations of quantum extreme learning machine," the authors contribute to the understanding of Quantum Reservoir Computing (QRC) and Quantum Extreme Learning Machines (QELM) frameworks. The main result is the relation between the training process and the reconstruction of the effective measurement characterizing given devices. The authors also provide a general mathematical model of QRC and QELM, which is concise and accessible.

The paper seems technically valid with a clear presentation. However, the interesting points of the outcomes in the manuscript are unclear. While this is the first paper to address the relationship between the performance of QRC/QELM with the conditional number (due to finite statistics), and the effect of multiple injections, the results are not surprising. Lacking the impact, I do not think the paper can influence thinking in the field. For example, the authors almost focus on QELM without considering the memory effects in QRC, which are more interesting perspectives. Extending the analysis to the QRC framework, such as investigating the memory effects, is necessary for considering publication in Communications Physics.

Several results need deeper discussion. For example, in the "multiple injections" section, typically, it is hard to prepare copies of the quantum state rho. Which motivation can be assumed if we need to reconstruct arbitrary functions of rho using multiple copies of rho? Furthermore, given a specific functon(not an arbitrary function) of rho, it will be more interesting to know how many copies of rho are needed for the reconstruction. It may relate to the problematic question of what kind of polynomial decomposition of rho is processed in the reservoir. In the classical context, we have a measure called Information Processing Capacity (Dambre, J., Verstraeten, D., Schrauwen, B. et al. Information Processing Capacity of Dynamical Systems. Sci Rep 2, 514) to evaluate this question. However, there are no similar things in the quantum context. Another interesting result that needs to be explained properly is the abrupt change in the MSE (Fig. 6) when the number of injections is larger than the degree of the polynomial of the target observable.

In summary, it is hard to be convinced about the impact of this paper. Without reliable support and explanation, I cannot tell that this research provides a substantial improvement or is applicable to significant physics problems. Therefore, the manuscript does not satisfy the quality to be published in Communications Physics.

Response to the Reviewers

We thank the Reviewers for their reading of our work and the constructive feedback. In the following we address their concerns point by point. For ease of perusal, relevant parts from the Reviewers' reports are reported in blue italic fonts, while our responses are in black fonts. Please, notice that the numbering of the figures has slightly changed in the revised manuscript: Fig. 2 now becomes Fig. 6 (in Appendix) while all other figures (except Fig. 1) have their label shifted by one (old Fig. 3 is now Fig. 2, etc.). Any reference to those plots has to be made with the old version of the manuscript.

Reviewer 1

Reviewer Point P 1.1 — *While the theory section is well detailed, the discussion of the results in Sections III an IV is rather limited, not covering/analyzing all the presented results. For instance Fig.4 and Fig.6(b) are not even discussed. Some specific recommendations to improve these sections are to*

- *clarify the meaning of an error of order* $1000(!)$ *in estimating* σ_x *(of order 1) in Fig.3(d). Maybe related to this (at least) singular result, the authors could also comment the effect of changing the random unitary (in this and the other figures).*

  Reply: We thank the Reviewer for the possibility to remark the reasons behind the disappointingly large values of the mean square error (MSE) associated with the reconstruction of σ_x. What it is happening in fig.3-(d) may be understood considering the different statistics used in estimating training and test probabilities, and the minimum number of outcomes necessary to yield a solution in the training phase. In the specific simulation reported in fig.3-(d), as we aim to reconstruct a qubit observable, we need at least 4 outcomes to perform the training. However, this is not sufficient to obtain small error results in the testing step because of the amplification – implemented by the map W – of the statistical error present in the vector of probabilities p. The amount of this amplification is well described by the condition number in Eq. (19). We have remarked this point in the revised version of the manuscript.

  It is also an interesting phenomenon that this amplification of error corresponding to 4 outcomes is only seen in fig.3-(d), where the training statistics is much larger than the statistics in the test phase, but not in fig.3-(b) and fig.3-(c). The reason can be tracked down to the fact that with infinitely accurate training datasets it is possible to retrieve the ideal linear map W, which works perfectly for perfectly estimated probabilities in the testing phase. However, this same map will also correspond to a larger amplification of error when there is noise in the test set. The consequence of this is that it is better to use a slightly inaccurate but more noise resilient map W for practical purposes, which will yield the results in fig.3-(b). In fact, a core reason behind our lengthy discussion about the statistics of training in section II-D, part of which was now moved to the appendix, was to explain this very same feature.

- *in Fig.3 explicitly state if this is the case of the scenario 1 and for which qudit size;*

Reply: Fig.3 is an example of the first scenario in which the qudit size is taken to be 2^5. The caption and main text have been updated to reflect this.

- *display the plot of the ideal case in Fig.3 to see the deviation from it;*

  Reply: As specified in the manuscript, in the ideal case the MSE is null when the number of outcomes is larger than 4. Therefore, a plot of such idealised situation would beunnecessary.

- *clarify if/that the x axis In Fig.3 refers to the output size; the names "number of samples" and "number of measurements" can be misleading. Does the latter correspond to m in sect II?*

  Reply: In Fig.3 the horizontal axis displays the number of possible outcomes, i.e. the dimensions of the alphabet Σ in Eq. (2), or the length of the vector $\mathbf{y}$ in Eq. (1) of the main text. We have changed the label in each panel of the figure to "number of outcomes".

- *in Fig.3 it should be clarified that all results are obtained for the test data (is it the case?); Are the training and test samples set to the size of the number of samples as indicated in the caption or is it just a problem of notation? Anyway provide the numbers.*

  Reply: In our study, the MSE is always evaluated in reference to the test dataset. We have taken $M_{\mathrm{train}} = 100$, $M_{\mathrm{test}} = 1000$ for all the plots as now more clearly stated in the revised manuscript. To avoid any inconsistencies with the notation, we additionally renamed $M_{\mathrm{train,test}}$ the number of training or testing points in the datasets, while N, as before, refers to the number of samples used to estimate the measurement probabilities.

- *comment on the dependence of these results when changing the Hamiltonian in Fig.3;*

  Reply: As specified in the caption of Fig. 3, chosing different observables does not significantly affect the behaviour of plots, and changing the random unitary has the same effect. In order to avoid the risk that such comment is overlooked, we have added it in the main text of Sec. III as well.

- *clarify what does exactly represent each dot in Fig. 4;*

  Reply: We are sorry that Fig. 4 lacked of the necessary degree of clarity. Each dot represents the results of simulations performed using a specific realization of a six-qubit reservoir, to which a condition number and a MSE are associated. As now explained more explicitly in the caption of the Figure, these two values are obtained by training the model with $M_{\mathrm{tr}} = 100$ states and testing it with $M_{\mathrm{test}} = 1000$ states, while probabilities are evaluated by simulating a finite statistics of $N = 10^4$ samples.

- *the authors actually do not draw any conclusion on Fig.4. Which is the conveyed message?*

  Reply: A new paragraph has been added to the main text to summarize and highlight the conclusions that can be drawn from Fig. 4. The paragraph reads: "Overall, as the degree of connectivity of the network increases, the performance of the reservoir and stability of the linear regression both improve. This is illustrated by the decrease in the MSE and the condition number."

- *Overall the results in Figs 3 and 4 are really hard to see and could be represented in larger sizes.*

  Reply: Thank you for bringing this to our attention. We have rearranged the plots in Figs 3 and 4 to improve their visibility.

- *Fig.5 does not really add too much insight (and is rather large);*

  Reply: Upon consideration, we have decided to remove Fig. 5 (b) altogether, which addressed QRC and was thus not central to our study.

- *clarify the meaning of the peak in 9 in Fig 6 (I could not understand it from the comment at the end of the section);*

  Reply: We regret the lack of clarity. As shown in Section IV.A, reconstructing a nonlinear functional of the state of a system through multiple injections is equivalent to the reconstruction of a linear function of a suitably large tensor product of the state with itself, see e.g. Eq. (27). When the number of injections is too high, the "effective" input space of Hermitian observables in Eq. (28) can become much larger than the dimension of the reservoir, thus potentially compromising the success of the reconstruction for some observables. This is explained in Section II.B, where it is stated that a necessary and sufficient condition for reconstructing Tr($\mathcal{O}\rho$) for every $\mathcal{O}$ is that the effective POVM $\tilde{\mu}$ should be informationally complete.

- *add a comment about the importance/relevance of Fig.6(b); comment on the steps arising for the exponential task as well as the "flatness" of the other one.*

  Reply: We thank the Referee for their suggestion. We have added the following paragraph at the end of section IV : "In fig.5-(b) [cf. old fig.6-(b)] we treat the case of non linear functionals of ρ. The performance achieved in approximating $\sqrt{1-\mathrm{Tr}(\rho^2)}$ is poor due to the slow convergence of the Taylor expansion of the functional. The step-like behavior that is evident in the MSE associated with the reconstruction of $\mathrm{Tr}(e^\rho)$, which is also present in the case of $\sqrt{1-\mathrm{Tr}(\rho^2)}$ although less evidently, can be explained by noticing that the trace of odd powers of ρ is a polynomial of the same degree of the previous even ones." A relatively easy way to see this is to work out the expression for $\mathrm{Tr}(\rho^2)$ and $\mathrm{Tr}(\rho^3)$ when the state is expanded in an orthonormal basis of traceless Hermitian operators. One quickly finds that the additional terms present in ρ^3 disappear computing the trace, and the same happens comparing ρ^{2n+1} with ρ^{2n}.

Reviewer Point P 1.2 — *The authors introduce a technical discussion about state pseudo-inverse and condition number in Section IID and Fig.2 but the main novel conclusions and implications of this analysis for QELM are not really clear, also looking at the growth with N and peak in Fig.3(a). Which is the use of this (peaked) figure of merit in order to assess the QELM performance? Given the unclear relevance of these findings, this long discussion on condition number could probably be shortened. As a comment, building on this discussion, do the authors consider any regularization procedure, as commonly considered in classical settings?*

Reply: We thank the Referee for their remark. We have moved part of Section IID and fig.2 to an appendix and shortened it for brevity and clarity. In our simulations no procedure of regularization has been applied.

Reviewer Point P 1.3 — *QRC is often mentioned but actually neither the theory framework, nor the results, cover this case that seems actually rather different. The recommendation is to remove unrelated sentences in abstract and introduction referring to results on QRC (anticipating results about "ELM or QRC" while only ELM is covered). Also the captions of table I and Fig.5 make a reference to QRC but do not address this case. Considering that QRC significantly departs from QELM the fact that it "can also be analysed along similar lines" remains to be proved. Finally, the statement about the memory in the conclusions is hard to follow looking at the contents of this work...*

Reply: We thank the Referee for raising this point. After careful consideration and in light of some of the comments made by Referee 3 too, we have decided to remove any substantive reference to quantum reservoir computing (QRC) from our study. While QRC may be useful for defining the broader context of our research, it is not relevant to the scope of our investigation. Instead, we will focus primarily on the QELM framework, which is the core of our work. By eliminating the main ties to QRC, we can better focus on the key aspects of our study and provide more concise and relevant results.

Minor

Reviewer Point P 1.4 — *Missing axis in Fig.2.*

Reply: The figure is now correct.

Reviewer Point P 1.5 — *In the introduction when referring to [24] maybe the authors meant "implementation of nonlinear input-output maps with memory" on a NISQ device.*

Reply: The sentence has been amended.

Reviewer Point P 1.6 — *The authors write that "characterisation of the class of tasks that are accomplishable through QRC- or QELM [...] is lacking", but this is actually addressed in most of the references they cite considering a broad spectrum of tasks. Also in the abstract they write about a lacking characterization of potential and limitations but actually these aspects have been addressed from different perspectives in several works, starting from the first one [10].*

Reply: The Referee raises a meaningful point. However, we would like to remark that our goal is to explore the tasks that can be solved by QELM *in general*, without focusing on specific problems (with the exception of the examples that we have used to illustrate our framework). In fact, our formalism points out which conditions should be met for perfect reconstruction in terms of the dimensions of the spaces of both reservoir and input.

Reviewer Point P 1.7 — *Notation: In Section C the observable O is defined from the target, while after eq.(5) the observable correspond to the QELM output. Should the notation be disambiguated? Also the use of vectors instead of single components in Sections B and C is not always clear.*

Reply: We have revised the notation so as to resolve any ambiguity: more precisely now $\mathcal{O}$ stands for the target observable, while $\tilde{\mathcal{O}}$ indicates the estimated one.

Reviewer Point P 1.8 — *In Section III the authors introduce 3 scenarios for QELMs. Some references to their use in the literature should be included. Furthermore the following examples could make explicit reference to corresponding relative scenario (1 or 2 or 3) to guide the reader, e.g. in Fig.3 or 4.*

Reply: The captions of figs. 3 and 4 have been modified to state more explicitly to which scenarios they correspond to. We furthermore added a sentence at the beginning of section III to provide some more context for the scenarios involved, and added a couple of references to refer to.

Reviewer 2

Reviewer Point P 2.1 — *First, the claims in the introduction part of manuscript are made very general and is not in agreement with the results shown in the rest of the manuscript. This must be addressed and need to be more accurate.*
For instance, see this paragraph: "In this paper, we make significant progress in this direction. First, we show that the problem of reconstructing features of a quantum state via an ELM-like setup can be viewed as a linear regression task on the measurement probabilities produced by a suitable positive operator valued measurement (POVM). The key observation is that the probability distribution corresponding to an arbitrary measurement of a quantum state is linear in the input density matrix. . . ."
In this paragraph, the claim is that the problem of reconstructing features of a quantum state via an ELM-like setup can be viewed as a linear regression task. This has been shown in several previous works. These previous works must be properly acknowledged. In fact, quantum reservoir networks have been successfully used for quantum state tomography, which is arguably the most complete for of feature reconstruction of quantum state. Given that these published works are directly related to the claims of the paper, I would suggest the authors to not claim these as their own original finding and cite the relevant papers in this paragraph. For instance, see these Refs [Phys. Rev. Lett. 127, 260401 (2021), and IEEE TNNLS, 32, 3148-3155 (2021)].
However, I would like to say that the study reported in the subsequent paragraph is important and original contribution. This paragraph can be more highlighted.

Reply: We have revised our introduction based on the Referee's feedback, incorporating their suggestions and recommended references.

Minor

Reviewer Point P 2.2 — *In Figure 2, the axes are not labelled.*

Reply: We have amended the figure to address this point.

Reviewer Point P 2.3 — *Figure 6 is missing in the manuscript.*

Reply: We are unsure about the nature of the problem pointed out by the Referee. We interpret it along the lines of what was pointed out by Referee 1, that is a lack of reference to Fig. 6 in the main text. We thus refer the Referee to our assessment of that criticism for a specific response.

Reviewer 3

Reviewer Point P 3.1 — *The paper seems technically valid with a clear presentation. However, the interesting points of the outcomes in the manuscript are unclear. While this is the first paper to address the relationship between the performance of QRC/QELM with the conditional number (due to finite statistics), and the effect of multiple injections, the results are not surprising. Lacking the impact, I do not think the paper can influence thinking in the field. For example, the authors almost focus on QELM without considering the memory effects in QRC, which are more interesting perspectives. Extending the analysis to the QRC framework, such as investigating the memory effects, is necessary for considering publication in Communications Physics.*

Reply: As mentioned in reply to Referee 1, after careful consideration, we have decided to remove any substantive reference to quantum reservoir computing (QRC) from our study. While QRC is useful for defining the broader context of our research, and is the natural candidate for further extending our results to the time-dependent setting, it is not strictly required to appreciate the main message of the manuscript. By eliminating any main ties to QRC, we can better focus on the key aspects of our study and provide more concise and relevant results.

However, we would like to point out that the Referee does not offer any argument to support their claim that our work lacks of impact. In a similarly subjective manner, their claim that memory effects in QRC surpass, in interest, the assessment of the performance of QELMs against the dimensions of the reservoir and the number of injections remains entirely unsupported and unjustified. While, obviously, the Referee is fully entitled to their opinion, it is difficult to address the criticism raised in this part of their report without any explicit and unbiased motivation for their statement. Needless to say, while we fully respect the opinion of the Referee, we strongly disagree in that we believe that our analysis and results pinpoint a novel connection that would be quite impactful for the development of experimental approaches to QELMs, as currently pursued in, among others, photonic platforms.

Reviewer Point P 3.2 — *Several results need deeper discussion. For example, in the "multiple injections" section, typically, it is hard to prepare copies of the quantum state rho. Which motivation can be assumed if we need to reconstruct arbitrary functions of rho using multiple copies of rho?*

Reply: We remark that our analysis addresses explicitly the problem of estimating non-linear functionals of the density matrix. It is a very well-acquired result, dating back to the seminal paper by A. Ekert et al. PRL **88**, 217901 (2002), that non-linear functionals require multiple copies of the state at hand, which aligns the resources required by our approach to well-established approaches that are not based on QELMs. One of the interesting messages of our manuscript is, in fact, that requiring multiple injections is not a limitation of our approach, but rather an intrinsic characteristic of QELMs.

Reviewer Point P 3.3 — *Furthermore, given a specific functon(not an arbitrary function) of rho, it will be more interesting to know how many copies of rho are needed for the reconstruction. It* *may relate to the problematic question of what kind of polynomial decomposition of rho is processed in the reservoir. In the classical context, we have a measure called Information Processing Capacity (Dambre, J., Verstraeten, D., Schrauwen, B. et al. Information Processing Capacity of Dynamical Systems. Sci Rep 2, 514) to evaluate this question. However, there are no similar things in the quantum context.*

Reply: We agree with the Referee on the relevance of the point that they raise. In fact, our study of the cases with multiple injections directly address precisely the kinds of functionals of the input state can be reconstructed. Specifically, as we discuss in the manuscript, one can reconstruct exactly polynomials of the input state of degree equal to the number of injections. This is not a limitation of our protocol, but rather an intrinsic characteristic of this class of QELM protocols. An explicit analytical relation between the number of injections and the performance of reconstruction did not emerge from our analysis, and is left for future studies.

Reviewer Point P 3.4 — *Another interesting result that needs to be explained properly is the abrupt change in the MSE (Fig. 6) when the number of injections is larger than the degree of the polynomial of the target observable.*

Reply: We respectfully point the Referee to the necessity of having informationally complete datasets. As pointed out already in our reply to Referee 1 and clearly discussed in the body of the manuscript, this establishes a link between the number of injections, the size of the reservoir and the behavior of the MSE.

In fact, a pivotal aspect emerging from our study is that the degree of achievable target polynomials is inextricably tied to the number of injections, meaning that k injections are needed to reconstruct polynomial targets of degree k. But at the same time, increasing the number of injections increases the dimension of the underlying state space, which becomes the space of symmetric tensor products of k states. We discuss these aspects in section IV-A.

REVIEWERS' COMMENTS:

Reviewer #1 (Remarks to the Author):

The authors have taken into account all suggestions, answered comments and questions, and accordingly modified the manuscript. All results/figures are now described and overall the presentation is improved. Some questions are left for future works but the scope of this work is broad enough. As two final notes, some arxiv references have now been published and can be updated, and |Sigma| vs. |mu tilde| (pg 7) should be disambiguated.

As mentioned in my previous report, the formal description of QELM in the language of channels and POVM and the description of the origins of non-linearity (beyond what is known for non-linear input encoding) is indeed interesting.
Therefore considering the paper content, the results are original with respect to other approaches, provide a consistent theoretical framework, and the formal treatment the authors present has the potential to have an impact and to be used and generalized in further treatments in quantum machine learning.
In conclusion, this work is interesting and novel, and in my opinion suitable for Communication Physics. I am therefore recommending it in the present form for publication.

Reviewer #2 (Remarks to the Author):

The authors have addressed all my concerns and updated the manuscript with the suggested changes. Thus, in my point of view, the manuscript is ready for publication.

Reviewer #3 (Remarks to the Author):

As mentioned in my previous report, the paper seems technically valid, but its impact is not convincing. I provided comments on further considerations for QRC, but I agree with the decision to remove the substantive reference to it. I do not have any biased opinion on the results of this paper. On the other hand, I suggested extending the analysis to more interesting results, such as the memory effects of QRC, in my opinion. I thought this would help strengthen the paper further. I might not have made such comments if the authors did not mention QRC with a solid story.

After reading the authors' responses to my comments and other referees' comments, I have decided to maintain my previous decision: "the manuscript does not satisfy the quality to be published in Communications Physics." While I see that the effect of the number of outcomes on the general performance is evident in Fig. 2, I do not find the core result of this paper convincing. For example, in the Introduction, the authors claim that "we discuss how suitable choices of effective POVMs can be used to improve overall performances," but this claim is unclear in the rest of the paper. Something may hinder me from understanding the core result in a concrete way. The paper has value in providing an overall analysis of the performance of QELM in standard scenarios rather than focusing on a specific Hamiltonian. However, it is difficult to determine whether the paper will influence thinking or be of interest to others in the field.

Finally, the authors should consider submitting the paper to other journals, such as Physical Review. With different criteria, this paper is publishable in such kind of journal.

Response to Referee #3's remarks.

> *As mentioned in my previous report, the paper seems technically valid, but its impact is not convincing. I provided comments on further considerations for QRC, but I agree with the decision to remove the substantive reference to it. I do not have any biased opinion on the results of this paper. On the other hand, I suggested extending the analysis to more interesting results, such as the memory effects of QRC, in my opinion. I thought this would help strengthen the paper further. I might not have made such comments if the authors did not mention QRC with a solid story.*

We made the choice of removing most references to QRC to be more in line with the majority of referees' remarks. We are aware Referee #3 considered some of these remarks interesting and even worth expanding. However, given the clash with some of the other Referees' opinion, we had to decide to leave the more in-depth discussion about QRCs for a dedicated work.

> *For example, in the Introduction, the authors claim that "we discuss how suitable choices of effective POVMs can be used to improve overall performances," but this claim is unclear in the rest of the paper. Something may hinder me from understanding the core result in a concrete way. The paper has value in providing an overall analysis of the performance of QELM in standard scenarios rather than focusing on a specific Hamiltonian. However, it is difficult to determine whether the paper will influence thinking or be of interest to others in the field.*

We agree with the referee about that specific sentence in the introduction being somewhat vague and unclear. We revised the manuscript to be more precise about the claims that are being made, adding the following sentence to the introduction:

> "*More generally, we show that the efficiency of QELMs is directly tied to the effective POVM summarizing both evolution and measurement. This puts the spotlight on the properties of this effective POVM, and on how these are the ones directly affecting performances.*"

We furthermore added the following remarks to the Conclusions:

> *"In particular, we showed that the effective POVM contains all of the information required to determine which observables can be estimated, and to what accuracy, as well as which kinds of effective POVMs, induced by different types of dynamics, result in different degrees of estimation accuracies.".*