nature portfolio

Peer Review File

Microwave signal processing using an analog quantum reservoir computer

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

In this article Senanian and collaborators have experimentally implemented reservoir computing on a quantum system composed of a microwave cavity coupled to a single transmon qubit, and they have applied it to three different machine learning tasks. This is the first experimental implementation on an analog quantum system. The authors have developed a scheme where an input is sent on resonance to the cavity, then the state of the qubit is measured, and finally the qubit is used to measure the parity of the cavity state, before everything is repeated for the following inputs. The whole sequence for all the input data is repeated multiple times and the feature vector is composed from the moments of the statistical distribution of the parity measurement results over different measurement shots.

This reservoir computing scheme is new and very interesting because it uses the backaction of the measurement, which is an intrinsically quantum phenomenon, to introduce nonlinearity necessary for the information processing by the neural network. The idea of using moments to capture the structure of the quantum fluctuations is new and important in the field of quantum machine learning, and thus I recommend the publication of this article in Nature Communications if they answer to my point-by-point comments and questions listed below.

1. In the main text an emphasis is put on the analog nature of the system, and the possibility to combine its roles as a quantum detector and as a quantum reservoir. On the other hand, quantum dynamics, the choice of the unitaries and quantum measurement backaction are only discussed in the Supplementary Material. Some discussion of their role would be important in the main text.

2. Are input and control pulses sent separately in time? How are they distinguished otherwise?

3. The moments are not well defined in the Figure 1. In Figure 1 (c) it should be written x_{n0}, x_{n1}, x_{n2} etc instead of x_0, x_1, x_2.... Also, they could be written in a recursive way in a few lines, mu_1, mu_2, mu_n, because all those products do not exist for

mu_1 for example.

4. M is not well defined in the main text. Is it the the number of inputs? Which also determines the number of unitaries, and the number of qubit measurements? Its value should be explicitly stated, especially as it is written that the trade-off concerning its value is important.

5. The role of reset is not clear enough, is there a reset after each input?

Questions and comments on the Supplementary Material:

1. In appendix D it is written that sometimes better performance is obtained with pseudo-inverse than with ADAM. That is surprising as ADAM is optimized to give better results. What is the explication for this?

2. Why do the authors use SoftMax?

3. It is not clear why the lowest MSE doesn't give the best accuracy.

Reviewer #2 (Remarks to the Author):

The authors present a nice study showing how to integrate a radio-frequency detector with a machine-learning algorithm using superconducting quantum bits. They demonstrate two classification tasks where they demonstrate processing the classical radio-frequency signals to determine the encoded data format or the statistical form of a noisy signal. I think this is an interesting result, but I find the presentation of the work is lacking in several regards. I encourage the authors to rework the manuscript based on the comments below.

- While the two tasks they study appear to take advantage of their approach, it is not clear why these are relevant for any application. Some justification of their chosen tasks is needed.

- As a detector of radio-frequency radiation, they need to provide standard metrics. What frequency does it operate, what is the bandwidth, what is the noise-equivalent power, etc. While some of the details can be found by carefully going through the supporting material, they should be given in the body text.

- From what I can see, they do not mention in the body text that they have to highly attenuated the detector signal as it goes from room-temperature to the base temp of the frig. This needs to be pointed out explicitly and the metrics requested above need to be given outside and at the base of the frig. Some comparison to other technologies should be made.

- Given the high attenuation of the signal, are there applications of their approach for measurement classical radio-frequency signals that are already at the base temperature of the frig?

- I find the discussion of the setup and the mathematical description to be lacking. They have cavity and qubit control signals. They should be given different names to avoid confusion. I don't see a discussion of the role of the (first) cavity control signal, which is summed with the detected signal. The diagram of the setup in the supplemental information, which shows 3 circuit elements, does not correspond to the first figure in the body text, where they show two circuit elements. In the supplemental setup figure, they should label each one of the circuits and they have to indicate the location of the signal they are measuring outside the frig. In the Hamiltonian, they say second and last terms, but they probably mean pairs of terms? Anyway, they do not define all the symbols (such as \sigma_x and \sigma_y, nor \Omega).

- They show that they can outperform a classical reservoir computer. The danger is that there is no proof that this is the classical reservoir computer. They are open to the issue that someone could come along and demonstrate a reservoir computer working with substantially fewer resources. For this reason, I don't think that this is a great comparison. I like better that they show decreased performance when they destroy entanglement in the system. They might consider dropping the comparison to the classical approach - it does not seem to be needed.

Reviewer #3 (Remarks to the Author):
This work demonstrates the use of a superconducting quantum circuit for analog quantum

reservoir computing, enabling the classification of microwave signals without discretization and the processing of ultra-low-power signals. The presented experimental implementation is an important advance for the field of quantum reservoir computing, which could merit publication in Nature Communications provided the authors address the following aspects.

1) The quantum reservoir computing literature discusses several measurement protocols to allow for continuous measurements. It is not clear to me how the current work relates to previous works, could the authors clarify?

2) I find some of the explanations of the system to be relatively involved unnecessarily. In particular, it is not clear to me what is the dimensionality of the feature vector. As described in Appendix D, the feature vector is defined to be R-dimensional but it remains quite abstract. In my opinion, it would help the reader to add a simple example of how to compute the value of R from the other parameter choices (order of moments used and number of shots). This would allow easier comparisons to classical reservoir computers. It would also be useful if the authors could show how many of the features are linearly independent for some particular cases (for instance in figure S17).

3) In my current reading of the manuscript, I understand the quantum coherences are a needed resource to increase the performance of the system and a major justification to claim that the experimental system is indeed working as a quantum reservoir computer. Is there a simple metric to relate the quantum advantage in terms of physical resources between a classical and a quantum reservoir?

4) In the context of reservoir computing, memory is a fundamental property. Here, the authors build feature vectors from correlations between nearest, next-nearest, and next-next-nearest measurements. Is this analogous to say that the memory of this QRC system is two steps? If not, could the authors comment on the memory of the system in physical units?

5) The current manuscript deals with microwave photons while there are some recent suggestions to implement QRC with optical photons, either with proof-of-principle

experiments or with numerical simulations of theoretically described implementations. The manuscript would benefit from a brief comparison to other photonic approaches to quantum reservoir computing.

Minor points:

- In the classification of radio-frequency (RF) communication modulation protocols, the protocols 32QAM and 32APSK tend to be misclassified. Did the authors estimate what would be needed to improve the accuracy in this task in terms of reservoir computing properties such as memory or nonlinearity?

- Is it possible to estimate the maximum information processing speed of this microwave QRC system in terms of bits per second?

- The sentence "Given the roughly 2×16 dimensions of Hilbert space used by our reservoir, this computational capacity is on the order of what would be expected for a large shot number." at the end of Appendix H merits additional explanations.

Reviewer #3 (Remarks on code availability):
I did not review the code.

Point-by-point response to the reviewers.

We would like to thank all the reviewers for their detailed and thorough reviews. We are glad that all the reviewers recognize the significance and relevance of our work. The reviewers raised several questions and made suggestions which have allowed us to significantly improve our manuscript.

Summary of main changes:
1. We have provided summaries of the design and motivation of the control signals in the main text, rather than referring the reader to the Appendix.
2. We have made clarifying edits to the main text regarding the output feature vector construction, as well as provided more details in the Appendix.
3. We have removed the section of the Appendix comparing our Quantum reservoir computer (QRC) to a classical reservoir computer (previously Appendix H).

In the following, we reply on a point-by-point basis. Reviewer comments are typeset in blue, our reply in black. Changes to the manuscript are typeset in green. Outside of grammatical corrections, all our changes are contained within the reply.

Reviewer 1:

In this article Senanian and collaborators have experimentally implemented reservoir computing on a quantum system composed of a microwave cavity coupled to a single transmon qubit, and they have applied it to three different machine learning tasks. This is the first experimental implementation on an analog quantum system. The authors have developed a scheme where an input is sent on resonance to the cavity, then the state of the qubit is measured, and finally the qubit is used to measure the parity of the cavity state, before everything is repeated for the following inputs. The whole sequence for all the input data is repeated multiple times and the feature vector is composed from the moments of the statistical distribution of the parity measurement results over different measurement shots.

This reservoir computing scheme is new and very interesting because it uses the backaction of the measurement, which is an intrinsically quantum phenomenon, to introduce nonlinearity necessary for the information processing by the neural network. The idea of using moments to capture the structure of the quantum fluctuations is new and important in the field of quantum machine learning, and thus I recommend the publication of this article in Nature Communications if they answer to my point-by-point comments and questions listed below.

We thank the reviewer for the concise summary of our main idea and for acknowledging its broader relevance in the field of quantum machine learning.

In the main text an emphasis is put on the analog nature of the system, and the possibility to combine its roles as a quantum detector and as a quantum reservoir. On the other hand, quantum dynamics, the choice of the unitaries and quantum measurement backaction are only discussed in the Supplementary Material. Some discussion of their role would be important in the main text.

Are input and control pulses sent separately in time? How are they distinguished otherwise?

We thank the reviewer for bringing this point of confusion to our attention. The cavity control pulses are sent before and after the input signal. The qubit control scheme features pulses before and after the input, as well as a full qubit rotation during the input signal. The other distinguishing feature between the input and control signals is that, while the control signals are fixed across all tasks, the input signals will vary from task to task (and within each task).

Action taken: To address the two comments above, we have rewritten the following paragraph in the introduction to include clearer explanations for the unitary control and measurement backaction:

"The oscillator and qubit control drives used in this paper realize a reservoir that consists of a series of entangling unitaries interleaved with qubit and oscillator measurements (Fig.1c). The analog input is sent resonantly to the oscillator and results in a time varying conditional displacement of the oscillator, which streams in concurrently with control drives. The cavity resonator hosting the oscillator mode has a frequency of 6 GHz and a 2-kHz linewidth. The combination of the input and control drives implements a unitary that encodes the input into the state of the oscillator and generate entanglement between the qubit and the oscillator. Following the unitary evolution, we perform a qubit measurement, and then a parity measurement of the oscillator state [38,39] (see Appendix C). The parity measurement projects the oscillator state into super-positions of either even or odd Fock states, a highly non-Gaussian measurement allowing one to sense changes in the photon number distribution. Additionally, the entangling dynamics between the measurements effectively implement a sequence of non-commuting measurements generating correlated measurement distributions that can then be used as complex output features...."

Additionally, we have added the following sentences to the results section discussing the details of the classification of time-independent signals:
"...The entangling conditional displacements are applied before and after the unknown input is fed into the system, and the qubit is rotated by π or $\pi/2$ pulses before, during, and after the input. Due to the qubit-state-dependent shift of the oscillator frequency by $-\chi$ (see first term of Eq 1) these qubit rotations serve to make the cavity sensitive to the input signal independent of the state of the qubit at the start of each round of input. Additionally, when combined with the conditional displacements on the oscillator, the control and input scheme impart a geometric area enclosed by the cavity trajectory onto the qubit, such that the phase of an unknown time-independent input signal can be extracted via a qubit measurement."

The moments are not well defined in the Figure 1. In Figure 1 (c) it should be written x_{n0}, x_{n1}, x_{n2} etc instead of x_0, x_1, x_2.... Also, they could be written in a recursive way in a few lines, mu_1, mu_2, mu_n, because all those products do not exist for mu_1 for example.

We thank the reviewer for pointing out another point of confusion. We have edited Figure 1 to reflect the above suggestion and have added clarifying text in the main article.

Action taken: We have added the following text to the introduction to help the reader understand how the output feature vector is constructed using examples: "For example, the first-order central moment μ_1 is a M-dimensional vector representing the average over all measured bitstrings, i.e. $\mu_1 = [\langle x_{n0}\rangle, \langle x_{n1}\rangle, \ldots]$, the second-order central moment μ_2 is the covariance matrix with elements $(\mu_2)_{ij} = \langle x_{ni}x_{nj}\rangle - \langle x_{ni}\rangle\langle x_{nj}\rangle$, and so on. Here, the expectation value is taken over the sample index n.

We have also made edits to Figure 1:
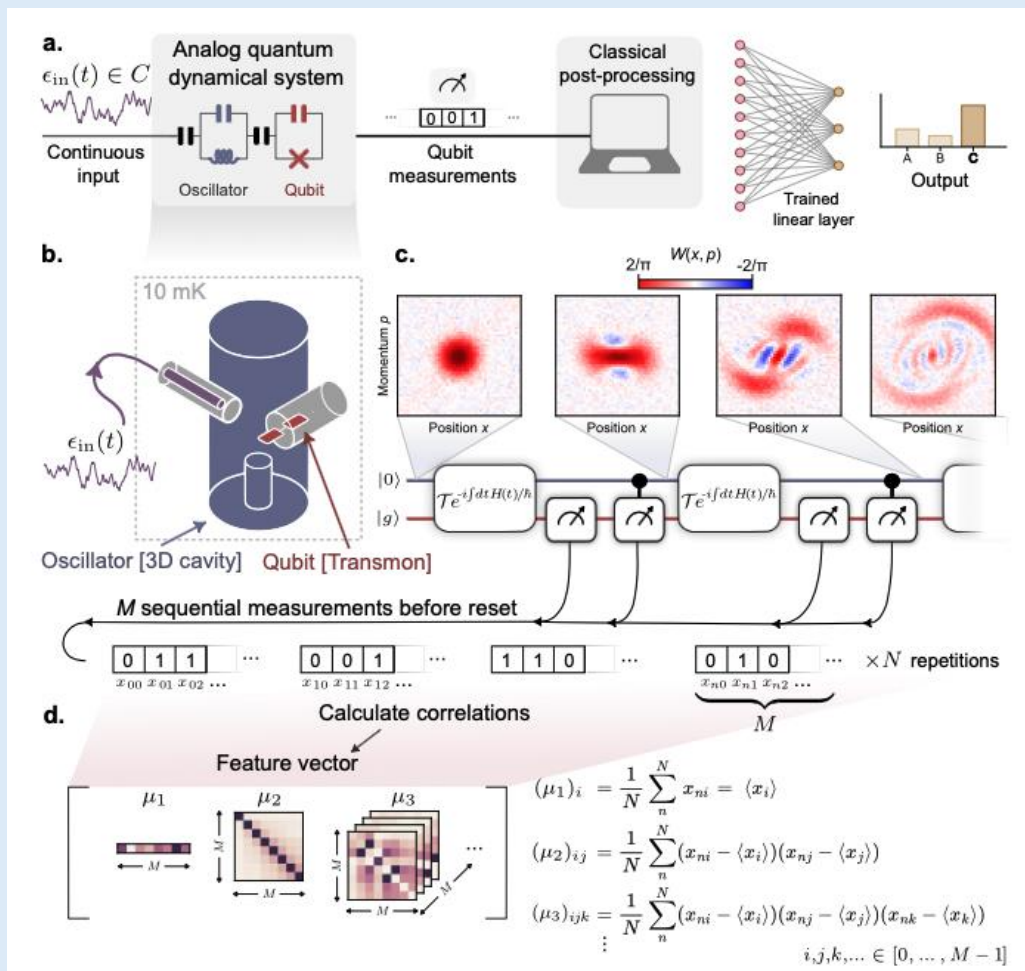
M is not well defined in the main text. Is it the number of inputs? Which also determines the number of unitaries, and the number of qubit measurements? Its value should be explicitly stated, especially as it is written that the trade-off concerning its value is important.

M is the number of measurements performed before the system is reset, and therefore corresponds the length of the bitstring samples used to construct our output features. There are two measurements for each round of input (and control unitaries), so the number of

unitaries applied before a reset is M/2. In all of the tasks presented in the paper, we use $M = 8$ measurements before reset. The role of reset is answered in the response below.

> **Action taken:** We have explicitly mentioned the number of input rounds we perform before the reset is applied: "Finally, after four rounds of applying the unitary and the qubit-oscillator measurements, we reset the system before repeating the scheme so that we may collect many samples of the measurement trajectory."
>
> We have included the following section at the end of the introduction highlighting both the number of measurements and the output feature dimension used for the tasks: "These truncated moments are then flattened and concatenated to construct our output feature vectors. In all, for the M = 8 measurements we use in this work, the resultant output feature vector size with this prescription is 94."

The role of reset is not clear enough, is there a reset after each input?

We thank the reviewer for this question. The reset is applied after four rounds of the input (and therefore eight measurements). This is done for two reasons:
1. The output features are constructed from the statistics of the distribution describing the measurement trajectory across the eight measurements. However, in each run, we are only given a sample of a possible measurement trajectory (the bitstrings in Figure 1). Therefore, in order to calculate the statistics of the underlying distributions, we require a reset of the system to uncorrelate sets of bitstrings.
2. While the time between measurements is much smaller than the qubit coherence time, we are still ultimately limited by the cavity lifetime. Applying the reset before substantial decoherence occurs in the cavity ensures that our system remains coherent and quantum throughout the entire run of the experiment.

> **Action taken:** We have added the following sentence in the introduction following the discussion of the unitary and measurements: "Finally, after four rounds of applying the unitary and the qubit-oscillator measurements, we reset the system before repeating the scheme so that we may collect many samples of the measurement trajectory. The reset rate, which occurs much faster than the decoherence rate of the cavity, additionally ensures that our system remains coherent."

In appendix D it is written that sometimes better performance is obtained with pseudo-inverse than with ADAM. That is surprising as ADAM is optimized to give better results. What is the explication for this?

We thank the reviewer for this excellent question. The reason for this is that, in our work, the pseudo-inverse method and ADAM are used to find linear layers at the minima of different loss functions. Since our classification method is fundamentally discrete - i.e. we identify the class simply based on whichever output vector entry is the largest - and since a zero-loss linear

layer is not generally achievable in practice, no continuous loss function can perfectly capture the highest-accuracy linear layer in its minima.

Consider an example where we assign labels of [1,0] and [0,1] to classes of a binary classification problem. Suppose we try to train our linear layer by minimizing the mean-squared error (MSE) loss between our output vector y and the correct labels. For this, the pseudo-inverse method is provably optimal for achieving the minimal MSE loss in the limit of the perturbation epsilon going to zero (up to numerical issues such as e.g. stability of the inversion) and ADAM would perform, at best, about as well. However, MSE loss does not perfectly correlate with classification accuracy - indeed we often got a better accuracy by keeping epsilon deliberately large. As an example of this imperfect correlation, the output [1.1, 1.0] for a label [0, 1] has a lower MSE loss than the output [0.5, 2.0], but would result in an incorrect classification.

As an alternative loss function, we try to train our linear layer by instead minimizing the MSE loss between the Softmax of our output vectors and the correct labels, rather than the output vectors directly. Now, there is no longer an analytic solution to minimize this loss function. Consequently, we must instead rely on a conventional optimizer such as ADAM. In this loss function, potentially very large separations between incorrect and correct labels, e.g. [-100, +1000] get mapped very closely to desired label [0, 1], whereas they would be seen as "high-loss" by the previous method. Indeed, in the previous example, the correctly classifying output [0.5, 2.0] now has a lower loss than the incorrectly classifying output of [1.1, 1.0]. However, even this loss function does not perfectly map on to classification accuracy. For example, consider a correct label [0, 0, 1] and outputs [3.466, 3.466, 3.526] Softmax -> [0.32, 0.32, 0.34] and [0, 6.234, 6.194] Softmax-> [0.001, 0.490, 0.509]. The latter yields an incorrect label but a lower loss than the former, which yields a correct label. Additionally, had we compared the output vectors directly as in the first method, the output with the correct prediction would have had the lower loss!

Why do the authors use SoftMax?

SoftMax is a conventional function to use for classifications in machine learning. Although it does not perfectly map onto inaccuracy of classification, it can be a bit more flexible the kinds of outputs that yield correct prediction compared to directly calculating MSE loss. Direct application of MSE only considers terms to be "low loss" if the output vectors directly are already close to the desired e.g. [0,1] label, whereas with softmax terms such as [-10, 2], [11, 31] can both be mapped closely to [0,1].

> **Action taken:** We have added the following sentences to Appendix D: "This approach used softmax, a popular choice for classifiers in neural networks [12] and back-propagation using the automatic differentiation package from PyTorch [13]. Training through back-propagation with an optimizer is now necessary since an exact analytic solution to minimize the loss no longer exists, unlike the case of the pseudo-inverse."

It is not clear why the lowest MSE doesn't give the best accuracy.

We thank the reviewer for this comment -- this is indeed somewhat counter-intuitive. This is because MSE loss *directly on the output vectors* does not map directly to inaccuracy. To use an example from above, the output [1.1, 1.0] for a label [0, 1] has a lower MSE loss than the output [0.5, 2.0], but would result in an incorrect classification. In cases where exact, zero MSE loss cannot be obtained, even the linear layer achieving the global minimum doesn't have to have the best accuracy. It's also important to note that we use MSE Loss in both methods - the difference is that in one case (pseudo-inverse) we calculate the MSE Loss of the output vectors directly, and in the other (ADAM) we pass the output vectors through SoftMax before calculating MSE Loss.

Action taken: We have added the following sentences to Appendix D: "The two approaches optimize the linear layer over different loss landscapes. This is because our classification method is fundamentally discrete - i.e. we identify the class simply based on whichever output vector entry is the largest - so no continuous loss function will capture the highest-accuracy linear layer in its minima in the general case."

Reviewer 2:

The authors present a nice study showing how to integrate a radio-frequency detector with a machine-learning algorithm using superconducting quantum bits. They demonstrate two classification tasks where they demonstrate processing the classical radio-frequency signals to determine the encoded data format or the statistical form of a noisy signal. I think this is an interesting result, but I find the presentation of the work is lacking in several regards. I encourage the authors to rework the manuscript based on the comments below.

We thank the reviewer for the concise summary and feedback on the manuscript presentation. In addition to the point-by-point responses below, we have made several edits for clarity and presentation in the text in response.

While the two tasks they study appear to take advantage of their approach, it is not clear why these are relevant for any application. Some justification of their chosen tasks is needed.

We thank the referee for this suggestion. We have added or modified sentences in the main text that justify the tasks. We did not choose the tasks because we anticipated that they would do well in our system, rather, we chose these tasks to probe specific features and capabilities of our system.

The first task of classifying the spiral arms in Fig. 2 demonstrates the ability of our quantum reservoir to process inputs nonlinearly, while also providing an illustrative and understandable example task. Additionally, as we outline in Appendix section G, for time-independent tasks such as this, our QRC is provably a universal function approximator. The second task of classifying radio frequency (RF) signals is a standard benchmarking task in the RF machine learning community and provides a rough point of comparison. The choice of this task is also broadly motivated by applications in signal identification and interception, though we do not wish to make the claim that our QRC could be deployed in a real-world setting for these today. We have modified sentences with any such insinuation. Finally, the third task of classifying noisy signals probes both the memory of our QRC and the ability to perform classification of very high-dimensional data beyond what has been achieved with alternative protocols. In this third task, we modify the output feature encoding to study and distill which correlations in the quantum fluctuations are important given correlations in the input signals.

As a detector of radio-frequency radiation, they need to provide standard metrics. What frequency does it operate, what is the bandwidth, what is the noise-equivalent power, etc. While some of the details can be found by carefully going through the supporting material, they should be given in the body text.

We thank the reviewer for this suggestion. If one were to operate our device as a traditional power detector, the operating frequency is 6 GHz, the bandwidth is about 2 kHz. The literature of superconducting qubit-cavity photon detectors using the same setup (e.g. Ref. [38]) does not include noise equivalent power (NEP) value, however we have estimated it be roughly 10^{-19}-10^{-20} W/$\sqrt{Hz}$. This figure is calculated from the equation,

$$NEP = \frac{\hbar \omega n_{min}}{\tau_{meas}} \frac{\kappa_i + \kappa_c}{\kappa_c} \frac{1}{\sqrt{\kappa_i + \kappa_c}}$$

Here, τ_{meas} is the minimum measurement time to resolve down to n_{min} photons, and κ_i (κ_c) is the internal (coupling) linewidth of our cavity. For the device here, τ_{meas} is about 1 microsecond, and n_{min} is roughly 0.05-0.1 photons. The first factor describes the incident power to our device, the second factor is the transmission coefficient for photons traveling at the base plate entering our cavity, which is roughly a factor of 40, and the last factor is a normalization due the measurement bandwidth. The reason for the transmission coefficient being so low is to thermalize the oscillator to the fridge rather than the transmission line. We did not explore the trade-off between stronger coupling and thermal population in the cavity. Instead, we decided to just set the coupling pin so that the cavity is weakly coupled to the transmission line, as is the standard practice.

The NEP, as calculated above, is limited primarily by the qubit readout fidelity and the transmission coefficient (i.e. if the cavity was critically coupled, the NEP would improve by a factor of 40). Other works, using an identical setup and measurement scheme, have entirely done away with limitations on qubit readout fidelity by using sophisticated post-processing techniques on repeated measurements [38], bringing their minimum detectable photon number down to ~10^-4, leading to another two orders of magnitude improvement in the NEP as calculated above.

The signals that we classified could have been composed of ~0.05 photons per run (or perhaps down to 10^-4 if the post-processing techniques of [38] were implemented) of the reservoir and have been *detectable* (e.g. for an example of such a detection, see Fig. S5). However, for the tasks we performed, the number of photons in the cavity is closer to 1-1.5 photon per run.

The reason for the difference in the number of photons needed for detection vs classification is that, in this work, we are doing a proof-of-concept primarily of processing on-device to perform classification. Having more photons increases the computational capacity of our device by leveraging a larger portion of the Hilbert space, allowing us to construct more complicated features in response to signals composed of a few photons. While the NEP figure is sufficient to characterize the ability of our device to detect microwave photons, it is not sufficient to characterize the ability of our device to perform the classification. The latter of which was found to be 10-20 times higher than the requirement for single-photon detection, but will generally depend on the task itself. Ultimately, we would like a system that is operated with the highest level of photon sensitivity while also performing the QRC processing. This is an important step for future work and will, as we mention in the Discussion, hopefully result in a system that can both receive signal photons with high efficiency and classify them.

Since there doesn't yet appear to be an established way to calculate the NEP for superconducting-cavity-based detectors in the literature, and the focus of our paper is on the processing rather than detection-sensitivity aspects, we would prefer to leave our back-of-the-envlope NEP calculation out of the present manuscript, as we are sufficiently unsure about the number derived above. Instead, inspired by the suggestion, we would like to include quantitative metrics that we are confident in report. **We have added the minimum number of photons per second that is needed to perform the classification task in a reasonable amount of shots**. We have also added the value of the frequency and bandwidth of our detector.

> **Action taken:** We have added the following to the introduction describing the minimum number of photons per microsecond needed to perform the classification: "The maximum amplitude of the input signal distribution $\max(|\varepsilon_{in}|)$ (i.e. the points in the spiral arms furthest away from the origin in Fig. 2b) was chosen such that the amount of displacement of the oscillator state initialized in vacuum would result in a coherent state with $\bar{n} = 0.3$ photons per round of input. This input amplitude was needed in order to perform the classification with sufficient accuracy in a reasonable amount of shots"
>
> We have also added the linewidth and frequency to the paragraph introducing our system: "The cavity resonator hosting the oscillator mode has a frequency of 6 GHz and a 2-kHz linewidth."
>
> Finally, we also described the coupling coefficient in the Appendix: "The cavity pin is set such that the oscillator mode is under-coupled to the transmission line by a factor of 40. While this reduces the transmission of photons incident on our device by a factor of 40, it keeps the oscillator state thermalized to the fridge rather than the transmission line."

From what I can see, they do not mention in the body text that they have to highly attenuated the detector signal as it goes from room-temperature to the base temp of the frig. This needs to be pointed out explicitly and the metrics requested above need to be given outside and at the base of the frig. Some comparison to other technologies should be made.

We thank the reviewer for raising this point -- indeed, this is an important clarification for our work. While the signals reaching our device comprise a handful of photons, the signals originate at room temperature and are highly attenuated. We have added text in both the introduction and the conclusion that highlight this point.

Action taken: We have modified a few sentences in the introduction to read: "The signals we classify are synthesized at room temperature and pass through 60 dB of attenuation before reaching our device. However, if instead one combines this analog quantum processing with a sensitive quantum detector of microwave radiation, as has already been previously demonstrated using superconducting circuits [34–38], then one can construct a system that achieves a quantum advantage in the task of combined sensing and signal processing of high temperature signals."

Additionally, we have modified a sentence in the conclusion to read: "While the signals classified in this work originate at room temperature and are highly attenuated before reaching the device, our experiments have shown that it is possible to accurately classify signals using a superconducting circuit even when there are only a few photons of signal in the superconducting circuit within any single run."

Given the high attenuation of the signal, are there applications of their approach for measurement classical radio-frequency signals that are already at the base temperature of the frig?

We thank the reviewer for this important question. Before directly answering the question regarding applications of our QRC for signals that originate at the base temperature, we wish to highlight a central point: while it is true one could find applications for signals that originate at the base temperature, it is not our belief that this is the only application for our QRC device. As other experiments have demonstrated, it is possible to obtain an advantage on classical signals with noise temperatures much larger than the device temperature [34,37]. Thus, with careful microwave engineering techniques, we envision that one could modify our experiment to perform machine learning processing on signals with a signal-to-noise ratio of order 1 originating at temperatures much higher than the device temperature.

On the question of applications of our scheme to signals originating at millikelvin temperatures, we envision using our QRC to potentially classify signals from other quantum devices. These could be classical signals, e.g. coherent states encoding some measurement result in a handful of photons, or quantum signals where a quantum device is directly interfacing with our QRC, and the task would be to classify a particular density matrix.

I find the discussion of the setup and the mathematical description to be lacking. They have cavity and qubit control signals. They should be given different names to avoid confusion. I

don't see a discussion of the role of the (first) cavity control signal, which is summed with the detected signal.

We thank the reviewer for pointing out this source of confusion. Another reviewer had a similar comment, and we have made edits to the introduction to help clarify the role of the control signals in the main text, rather than refer the reader to the Appendix.

Action taken: In response to the clarification of the role of the control signals, we have added the following text describing the choice of control drives: "The unitary encoding the input displacement is complimented by control drives that entangle the qubit and cavity via a couple of conditional displacements [42] and qubit rotations (Fig. 2a). The entangling conditional displacements are applied before and after the unknown input is fed into the system, and the qubit is rotated by π or $\pi/2$ pulses before, during, and after the input. Due to the state-dependent shift of the cavity frequency by $-\chi$ (see first term of Eq 1) these qubit rotations serve to make the cavity sensitive to the input signal independent of the state of the qubit at the start of the protocol. Additionally, when combined with the conditional displacements on the cavity, the control and input scheme effectively impart a geometric area enclosed by the cavity trajectory onto the qubit, such that the phase of an unknown time-independent input signal can be extracted via a qubit measurement."

The diagram of the setup in the supplemental information, which shows 3 circuit elements, does not correspond to the first figure in the body text, where they show two circuit elements. In the supplemental setup figure, they should label each one of the circuits and they have to indicate the location of the signal they are measuring outside the frig.

We thank the referee for pointing out these important omissions. We have made edits that label each circuit element and labeled the origin of the input signals. Additionally, we have made an edit in the caption of Figure 1 to explicitly mention that the readout resonator is not shown for clarity in the main text.

Action taken: We have also made edits to Supplemental wiring diagram with clear labels for each of the components of our QRC device. We have also labeled the origin of our input signals (which are synthesized in the same way as our cavity control signals):
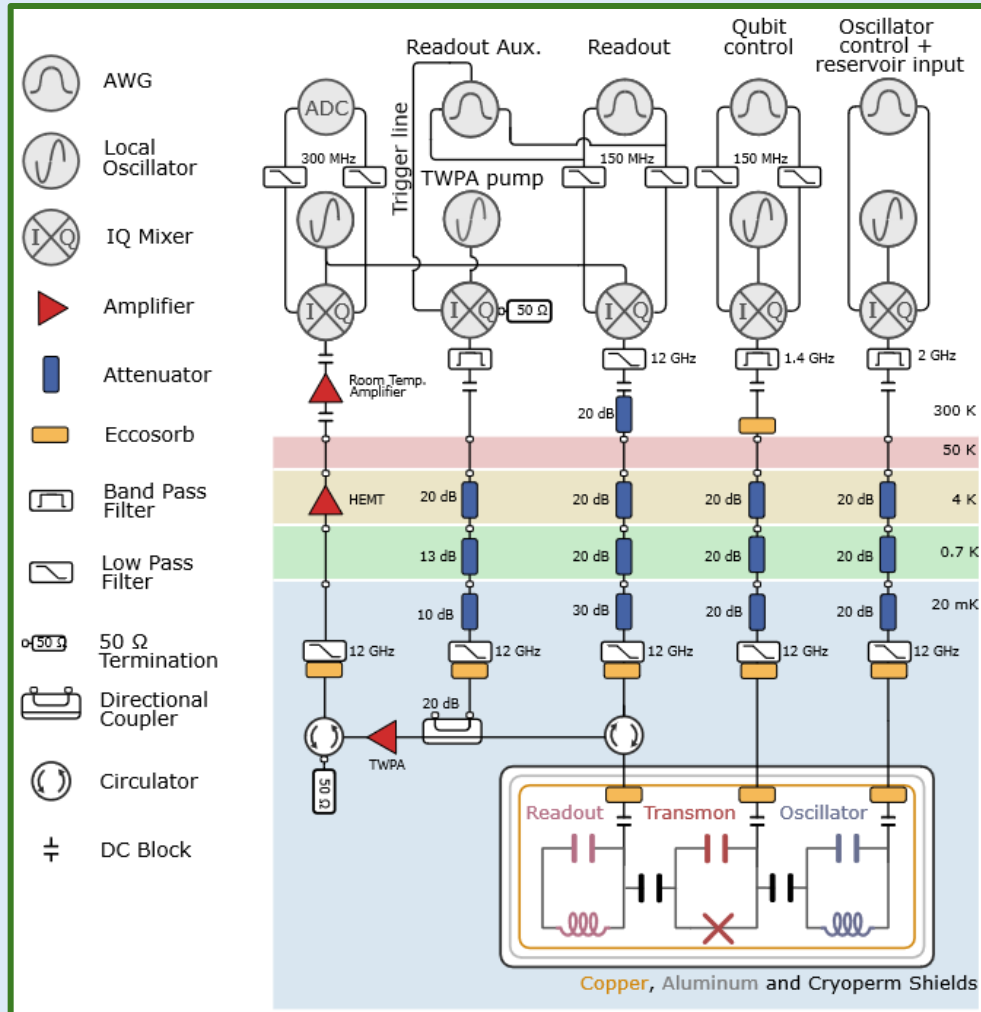
Additionally, in the caption for Figure 1 in the main text (not the figure above), we have explicitly stated that the readout resonator has been omitted in the schematic diagram: "The signals interface directly with the qubit-oscillator system, composed of a 3D aluminum cavity (blue) hosting a transmon qubit (red) and a readout resonator (omitted for clarity)."

In the Hamiltonian, they say second and last terms, but they probably mean pairs of terms? Anyway, they do not define all the symbols (such as \sigma_x and \sigma_y, nor \Omega).

We thank the reviewer for pointing out this mistake. We edited the text to describe all symbols defined in the Hamiltonian.

Action taken: We have edited the paragraph following the introduction of the Hamiltonian to include a definition of all terms. For brevity, the qubit operators are now explicitly written in terms of the Hilbert space, rather than the pauli matrices:

"Our quantum reservoir, composed of a cavity resonator coupled to a transmon (Fig. 1b), can be modeled with the following qubit-oscillator Hamiltonian in the rotating-frame

$$H = -\chi |e\rangle\langle e| a^\dagger a + \epsilon(t) a^\dagger + \Omega(t) |e\rangle\langle g| + H.c.$$

Where $|g\rangle$ and $|e\rangle$ define the qubit subspace of the transmon, a is the photon annihilation operator of the cavity mode, and χ is the nonlinear interaction strength (see Appendix B for details). The third term of Eq 1 describes the unitary control of the qubit from a time-dependent drive $\Omega(t)$, and the second term describes both the encoding of the input data $\epsilon_{in}(t)$ and unitary control of the oscillator mode, i.e., $\epsilon(t) = \epsilon_{in}(t) + \epsilon_{control}(t)$..."

They show that they can outperform a classical reservoir computer. The danger is that there is no proof that this is the classical reservoir computer. They are open to the issue that someone could come along and demonstrate a reservoir computer working with substantially fewer resources. For this reason, I don't think that this is a great comparison. I like better that they show decreased performance when they destroy entanglement in the system. They might consider dropping the comparison to the classical approach - it does not seem to be needed.

This point is absolutely true, and one we considered deeply while preparing the manuscript. Indeed, it is hard to argue for an objective classicalized reservoir. We have removed the Appendix section that compares our QRC with a classicalized version of the same device, however, we wish to keep the comparison with digital reservoir computers. While the construction of the former is subjective and from our own attempts at providing a fair comparison, the latter is a standard architecture for digital reservoir computers in the field.

Action taken: We have removed the Appendix section comparing the performance with a classicalized version of our quantum reservoir computer.

Reviewer 3:

This work demonstrates the use of a superconducting quantum circuit for analog quantum reservoir computing, enabling the classification of microwave signals without discretization and the processing of ultra-low-power signals. The presented experimental implementation is an important advance for the field of quantum reservoir computing, which could merit publication in Nature Communications provided the authors address the following aspects.

We thank the reviewer for the concise summary and for acknowledging its application to the field of quantum reservoir computing broadly.

The quantum reservoir computing literature discusses several measurement protocols to allow for continuous measurements. It is not clear to me how the current work relates to previous works, could the authors clarify?

We thank the reviewer for the question. Indeed, previous proposals for quantum reservoir computing include protocols with continuous measurements such as Refs. [22,23], where, like our work, the measurements contribute to the dynamics. In our experiments, the measurements employed are projective measurements, the backaction of which induce highly non-classical states within the reservoir dynamics. This is in contrast to the weak measurements employed in the proposals referenced previously. Additionally, in these experiments, we do not average over measurement outcomes, which would amount to only considering the first-order central moment. We build output feature vectors from higher-order moments in time, directly capturing the correlations induced by the act of measurements on a quantum system. Additionally, in this work, we perform machine learning classifications of analog signals, including very high-dimensional continuous-time signals, as opposed to parameter estimation of a parametrized signal.

> **Action taken:** We have added a sentence in the discussion to highlight the benefits of our platform and measurement protocol: "The superconducting circuits platform not only allows us to leverage projective non-demolition (QND) non-Gaussian measurements to generate correlated output features, but is also well-matched to process microwave signals that can generally be continuous in time."

I find some of the explanations of the system to be relatively involved unnecessarily. In particular, it is not clear to me what is the dimensionality of the feature vector. As described in Appendix D, the feature vector is defined to be R-dimensional but it remains quite abstract. In my opinion, it would help the reader to add a simple example of how to compute the value of R from the other parameter choices (order of moments used and number of shots). This would allow easier comparisons to classical reservoir computers. It would also be useful if the authors could show how many of the features are linearly independent for some particular cases (for instance in figure S17).

We thank the reviewer for pointing out this source of confusion. The dimensionality of the output feature vector depends on the order of the correlator, as well as how far apart we choose to keep the correlations. E.g., if we only keep up to second-order correlations among

nearest-neighbor and next-nearest-neighbors, this will result in a feature vector much smaller than if we also included higher-order correlators or looked at correlations among measurements further apart. The dimension of the output feature vector does not depend on the number of shots, however. The number of shots increases our ability to accurately estimate these correlators. For the tasks presented in this work, we use output feature vector size we use for our tasks is ninety-four.

We have edited the main text with explicit mentions of the feature vector dimension. Additionally, we have added edited the Appendix section discussing the construction of the output feature vector with examples, as well as explicit mentions of the output feature size for different order of moments/locality of the moments.

Regarding the comment on linear dependence of the output features, this will be task dependent, however, we do not expect to see a large amount of independence for many tasks. For example, in Figure S11 and S12 we see that the output features have a significant amount of redundancy for the different classes, despite the efforts to distill the output features to a physically significant encoding. Nonetheless, we think it is an interesting comment, and designing the output feature vectors is paramount to designing a good reservoir computer.

Action taken: In the introduction, we have edited the text to reflect the output feature vector construction as well as explicitly mentioned the dimension size: "For example, the first-order central moment μ_1 is a M-dimensional vector representing the average over all measured bitstrings, i.e. $\mu_1 = [\langle x_{n0}\rangle, \langle x_{n1}\rangle, \ldots]$, the second-order central moment μ_2 is the covariance matrix with elements $(\mu_2)_{ij} = \langle x_{ni}x_{nj}\rangle - \langle x_{ni}\rangle\langle x_{nj}\rangle$, and so on. Here, the expectation value is taken over the sample index n. This approach, inspired by Ref. [17], has the benefit of leveraging the hierarchy of noise in the central moments, while capturing the essential correlations in the dynamics to achieve high accuracy even in the few-sample regime. Furthermore, the output feature vector dimension only scales polynomially with the number of measurements, where the highest polynomial power is given by the order of the highest central moment, which we restrict to 3 for all tasks in this work. Finally, given finite memory in our reservoir, we further restrict the output vector by choosing to only calculate correlations between measurements at most 3 measurements apart. These truncated moments are then flattened and concatenated to construct our output feature vectors. In all, for the M = 8 measurements we use in this work, the resultant output feature vector size with this prescription is 94."

Additionally, in the Appendix, we have edited a paragraph to be more transparent in how we construct the output feature vectors:
"We denote these appended feature vectors as \vec{μ}$_{\leq p}$ for feature vectors containing up to p central moments, e.g.

$$\vec{\mu}_{\leq 2} = [\vec{\mu}_1, \mu_2]$$
$$= [\langle x_0\rangle, \langle x_1\rangle, \langle x_2\rangle, \ldots, \langle x_0 x_1\rangle - \langle x_0\rangle\langle x_1\rangle, \langle x_0 x_2\rangle - \langle x_0\rangle\langle x_2\rangle, \ldots, \langle x_1 x_2\rangle - \langle x_1\rangle\langle x_2\rangle, \ldots]$$

is a feature vector constructed from appending the flattened covariance to the mean. The first-order moment here is a vector to denote that we take the mean over repetitions of different measurements, whereas the covariance is a matrix and thus is not denoted as a vector. Additionally, we only take the independent degrees of freedom of the symmetric covariance matrix equivalent to discarding one of the following redundant elements $\langle x_i x_j\rangle$ and $\langle x_j x_i\rangle$ for some integers i,j. In general, for arbitrary moments, the number of independent components for M measurements is ${M+p-1}\choose{p}$, where p is the order of the moment. For up to third-order central moments of $M = 8$, this gives a total output feature dimension of dim(\vec{μ}$_{\leq 3}$) = 8 + 36 + 120 = 164. This output dimension for the results presented in the main text is further reduced as discussed in the following paragraphs below."

In my current reading of the manuscript, I understand the quantum coherences are a needed resource to increase the performance of the system and a major justification to claim that the experimental system is indeed working as a quantum reservoir computer. Is there a simple metric to relate the quantum advantage in terms of physical resources between a classical and a quantum reservoir?

We thank the reviewer for this excellent question. Indeed, this is a question we spent much time thinking about, but it is not an easy question to answer. The central issue is coming up

with a set of control drives that work across several different tasks. This was the central question that motivated the design for the control signals for both the quantum reservoir computer presented in the main text, as well as the classical reservoir computer presented in the Appendix. What was clear was that the set of control drives that seem to generalize well in the quantum case did not generalize in the classical case, and a separate design was needed in the classical reservoir computer, and this is what was pursued. However, as another reviewer pointed out, there is no guarantee that the control drives we produced are the optimal ones (for either the quantum or the classical case). Thus, since we agree with the other reviewer, we have decided to remove the section altogether.

> **Action taken:** We have removed the Appendix section comparing the performance with a classicalized version of our quantum reservoir computer.

In the context of reservoir computing, memory is a fundamental property. Here, the authors build feature vectors from correlations between nearest, next-nearest, and next-next-nearest measurements. Is this analogous to say that the memory of this QRC system is two steps? If not, could the authors comment on the memory of the system in physical units?

We thank the reviewer for the good question. We saw that for the tasks studied in this work, when we truncated the output feature vector to nearest, next-nearest, and next-next-nearest measurements, we saturated the accuracy. On the other hand, the correlations can last longer than the three sets of measurements, as indicated by the central moments plotted for the spiral classification task in Figure S11 and S12, though these longer-range correlations do not strongly contribute to distinguishing the input datasets between the two tasks.

The memory of our device is limited by the coherence of the cavity (100 microseconds for us here), or the duration before resetting the device (roughly 15 for us here) – whichever is shorter. Whether or not this memory leads to an improved accuracy will generally be task dependent, as noted in Figure 4 of the main text.

Finally, a note on the main motivation for keeping the reset time at roughly 15 us rather than nearly the entire cavity coherence time. Increasing the reset time in our scheme would lead to the added cost of having more measurements. More measurements mean more measurement trajectories, and therefore more sampling noise in estimating the correlators. Although our feature vector construction only requires several samples that grows polynomially in the number of measurements (as opposed to an exponential dependence as required by using the full measurement distribution), this polynomial cost was still significant, so we decided to keep the number of measurements to eight.

The current manuscript deals with microwave photons while there are some recent suggestions to implement QRC with optical photons, either with proof-of-principle experiments or with numerical simulations of theoretically described implementations. The manuscript would benefit from a brief comparison to other photonic approaches to quantum reservoir computing.

We thank the referee for the suggestion. The primary technological advantage with microwave quantum systems that we leverage in this work is the accessibility to highly non-Gaussian

QND measurements. While optical approaches have been proposed using non-Gaussian measurements such as photon-counting, the realization of such measurements in a fashion that is QND has not been demonstrated in the optical domain to the best of our knowledge. It is the combination of the QNDness and the non-Gaussianity of our measurements that allow us to generate states with Wigner negativity through measurement back-action.

> **Action taken:** We have added the following sentence in the discussion highlighting the advantage of the microwave platform: "The superconducting circuits platform not only allows us to leverage projective non-demolition (QND) non-Gaussian measurements to generate correlated output features but is also well-matched to process microwave signals that can generally be continuous in time."

In the classification of radio-frequency (RF) communication modulation protocols, the protocols 32QAM and 32APSK tend to be misclassified. Did the authors estimate what would be needed to improve the accuracy in this task in terms of reservoir computing properties such as memory or nonlinearity?

We believe the easiest way to reach > 99% on this task is to increase the number of measurements and the number of samples. For the current accuracy curve plotted in Fig. 3 in the main text, the accuracy has not yet saturated with the number of shots. We believe another factor of 5 shots would saturate the accuracy at > 99%. Indeed, adding more nonlinear elements would additionally increase the complexity of our system and likely lead to higher classification accuracy as was shown in simulations of multi-qubit systems in Appendix Fig. S18.

Is it possible to estimate the maximum information processing speed of this microwave QRC system in terms of bits per second?

We thank the referee for the interesting question. We think this question may be more relevant for a signal processing task that might involve decoding a modulated signal, or perhaps a task that discretely encoding a signal and performs a set of discrete gates. However, we feel, that ascribing bits per second to an analog system processing analog tasks becomes difficult, especially in a machine learning context where the computation is a classification of different analog signals.

The sentence "Given the roughly 2×16 dimensions of Hilbert space used by our reservoir, this computational capacity is on the order of what would be expected for a large shot number." at the end of Appendix H merits additional explanations.

Indeed, we have added a clarification to this point. In the conventional view of reservoirs, we make our data linearly separable by sending it into a higher dimensional space where such separability is possible. The dimensionality of our Hilbert space, given by the two dimensions of the qubit and the approximately 16 occupied levels of our storage resonator, limits the degrees of freedom we have available in encoding our data as the final, measured state. Consequently, we use this total Hilbert space dimension as a proxy for the quantum resources used in terms of reservoir dimensionality.

Action taken: We have added the following clarification in the Appendix regarding the effective dimensionality "In the conventional view of reservoirs, the data must be sent into into a higher dimensional space where linear separability becomes possible [10]. The dimensionality of our Hilbert space, given by the two dimensions of the qubit and the approximately 16 occupied levels of our storage resonator, limits the complex degrees of freedom we have available in encoding our data as the final, measured state. Consequently, we use this total Hilbert space dimension (times two due to complex amplitudes) as a proxy for the quantum resources used in terms of reservoir dimensionality. Given the roughly 2 * 16 dimensions of Hilbert space used by our reservoir, achieving *at least* the computational capacity of a 64-dimensional LESN reservoir is on the order of what would be expected for a large shot number. Indeed, the point of this comparison is to demonstrate that the computations performed by our reservoir cannot be trivially replicated by a digital reservoir with fewer resources with the same performance."

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

The authors have addressed all my comments. I thus support the publication of their work in Nature Communications. A single additional comment:

In Supplementary Material there is a typo in Eq.(B1), it should be $\epsilon^\ast (t) a + \epsilon (t) a^\dagger$. In the code it is ok.

Reviewer #2 (Remarks to the Author):

The authors have done an excellent job of addressing the concerns of all referees and the manuscript is now ready for publication.

Reviewer #3 (Remarks to the Author):

The authors have partly addressed my comments from my first report. First of all, I would like to thank them for clarifying the size of the output feature vector and the dimensionality of the quantum reservoir. However, some of my previous questions remain unanswered. Since I believe there was a slight misunderstanding, likely due to our different backgrounds, I would like to rephrase some of my previous comments more directly:

1) The memory of reservoir computing systems can be characterized by the so-called short-term memory capacity functon(see definition in https://www.ai.rug.nl/minds/uploads/STMEchoStatesTechRep.pdf, eq. 15). Did the authors characterize the memory capacity of their quantum reservoir computing system?

2) I would like to bring another work on photonic quantum reservoir computing to the attention of the authors, namely Spagnolo et al., Nat. Photon. 16, 318–323 (2022), which covers an experimental implementation that can handle the same type of tasks discussed in their manuscript. I would like to ask the authors for their opinion on the scalability and applicability of their approach in the context of the current literature (for instance, the

paper by Spagnolo et al. mentioned above).

3) In my previous report, I asked the authors about the processing speed of the system, but I did not receive an answer. Since the system is an analog one, could the authors report the input rate in inputs per second or in analog samples per second?

We thank the reviewer for the feedback and the recommendation.

We thank the referee for pointing this out. This is indeed incorrectly written, however we believe the correction is different than what is stated above: The Hamiltonian in Eq. B1 is written in the lab frame, whereas a term like $\varepsilon^*(t)\, a + \varepsilon(t)\, a^\dagger$ would only be valid in the rotating frame, since after all, our microwave signal generators can only output real-valued signals. However, it is indeed a typo to label the displacement drive as $\varepsilon(t)$ in both the rotating frame and the lab frame, since these will be quantitatively different. **We have relabeled the displacement drive terms in the lab frame (Eq. B1).**

To clarify the above, if the real-valued signal outputted from our signal generators is $\xi(t)$, then the rotating-frame displacement drivers $\varepsilon(t)$ are related to the lab frame drives as $\xi(t) = \varepsilon(t)\, e^{i\omega t} + \varepsilon^*(t)\, e^{-i\omega t}$ for some carrier frequency ω. The drive term in the lab-frame Hamiltonian is then:

$$H_d/\hbar = \xi(t)(a + a^\dagger)$$

$$H_d/\hbar = (\varepsilon(t)\, e^{i\omega t} + \varepsilon^*(t)\, e^{-i\omega t})(a + a^\dagger)$$

If we set ω to the cavity frequency and move into the rotating frame, we can expand the above factor. With this expanded, we will first have the following two terms

$$\varepsilon^*(t)\, a + \varepsilon(t)\, a^\dagger$$

Which are the same as the ones pointed out, and the ones in the Hamiltonian in the main text. In addition, we will have another two terms

$$\varepsilon^*(t)\, e^{2i\omega t}\, a^\dagger + \varepsilon(t)\, e^{-2i\omega t} a$$

However, these are oscillating at twice the cavity frequency (order of a few GHz away). Under the rotating wave approximation, we can safely drop these terms to obtain the displacement drive term found in the Hamiltonian of the main text.

Action taken: We have relabeled the symbol denoting the storage drive term in the lab frame as ξ , instead of incorrectly using the same label for the displacement drive term in the rotating frame

$$H/\hbar = \omega_q q^\dagger q + \omega_a a^\dagger a - \chi q^\dagger q a^\dagger a - \chi' q^\dagger q a^{\dagger 2} a^2 - K_q q^{\dagger 2} q^2 - K a^{\dagger 2} a^2 + \Xi(t)(q + q^\dagger) + \xi(t)(a + a^\dagger)$$

We have also described the relationship between the two in the text below Eq. B1: "The last two terms describe the qubit and oscillator drives in the lab frame. The lab-frame drives are related to the rotating-frame drives in Eq. 1 via $\Xi(t) = \Omega(t)e^{i\omega_q t} + $ H.c. and $\xi(t) = \varepsilon(t)e^{i\omega_a t} + $ H.c."

Reviewer #2 (Remarks to the Author):

The authors have done an excellent job of addressing the concerns of all referees and the manuscript is now ready for publication.

We thank the reviewer for the feedback and for the recommendation.

Reviewer #3 (Remarks to the Author):

The authors have partly addressed my comments from my first report. First of all, I would like to thank them for clarifying the size of the output feature vector and the dimensionality of the quantum reservoir. However, some of my previous questions remain unanswered. Since I believe there was a slight misunderstanding, likely due to our different backgrounds, I would like to rephrase some of my previous comments more directly:

1) The memory of reservoir computing systems can be characterized by the so-called short-term memory capacity functon(see definition in https://www.ai.rug.nl/minds/uploads/STMEchoStatesTechRep.pdf, eq. 15). Did the authors characterize the memory capacity of their quantum reservoir computing system?

We thank the referee for this suggestion. We have numerically evaluated the memory capacity of our quantum reservoir computing system based on simulations using the Eq. 15 from the above reference. Before discussing the results, we would like to point out a couple of caveats that makes the above definition of memory capacity difficult to implement in our analog implementation:

The reference above characterizes the short-term memory capacity of a digital reservoir computer by sampling a translated time-series as input, and evaluating a measure of similarity between the output and the sampled input for some translation k. This input sample is a single number taken by evaluating a time-domain signal at time t, and a feature vector is constructed for each input sample at t. After a dot-product is taken with the feature vector and a weight vector, a single output number is evaluated at time t. This procedure results in an *input and*

output vector as a function of time. The similarity between these two vectors as a function of the translation k is then evaluated for the memory capacity.

In our implementation of a quantum reservoir computer, inputs are received as time-domain analog signals, and output feature vectors are constructed from correlators across many measurements which contribute to the dynamics. This has two important consequences:

1. The closest correspondence to a single number input to our quantum reservoir would be a time-domain signal with a fixed amplitude and finite duration. For an example of this, see the spiral classification task in Fig. 2 of the main text.
2. To achieve a singular feature vector for a single input, this input must be repeated several times. This is a direct consequence of constructing our output feature vectors from correlators of measurements, which occur intermingled with the repeated inputs.

The consequence of points 1 and 2 above is that the smallest increment of translation k that fits the definition of capacity in the above reference is a duration of about 4 microseconds, and about 8 measurements. Therefore, given that the we experimentally found that covariances vanish for measurements about 3 or 4 apart (as noted in Appendix D, also see Fig. 4e), which is smaller than the smallest translation that fits the above definition for digital reservoirs, we do not expect the memory capacity to be finite for k > 1 using the above definition. This is also precisely what we observe in our simulations using the above adaptation of memory capacity to our analog system, plotted below.
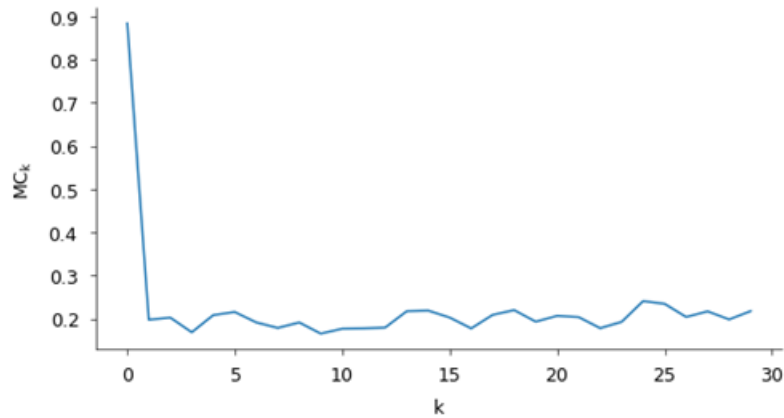

Figure 1: Memory capacity as a function of delay

To summarize: by adapting our analog system to receive digital inputs so that the memory capacity defined by the above reference may be evaluated, we strip away much of the ability of our system to store information. On the other hand, the output features of our quantum reservoir are constructed directly by calculating correlations across measurements (in time), which would all vanish if there was no memory in the system. Therefore**, since we find finite correlations across measurements, we know that there is some small but finite memory in our system**, which the above procedure cannot capture.

We thank the referee for making us aware of this work – it is indeed very relevant. In it, Spagnolo *et al.* fabricate a quantum memristor and numerically investigate its use as a quantum reservoir computer. Similar to our work, one of the tasks performed with the quantum reservoir computer is to classify time-dependent input signals using repeated measurements. In Spagnolo et al., the repeated non-QND measurements contribute to the dynamics by measurement feedback that influences the unitary performed in subsequent passes. This feedback is implemented such that short-term memory is enforced in the dynamics of the quantum reservoir.

A key difference between this work and ours is that in our work, the measurements are QND and are repeated before the state is decohered. Consequently, the measurements directly imprint quantum features on the state via measurement-backaction, as is experimentally demonstrated in the Wigner functions in Figure 1 of the main text. Fast and QND measurements are a unique advantage to superconducting-qubit systems. Another distinction is that in our work, the output feature size grows with the number of measurements, allowing us to efficiently construct high-dimensional and correlated features without the use of many physical resources. The quantum reservoir numerically studied in Spagnolo et al. constructed feature vectors from measurements obtained only in the last pass through the reservoir, though the vectors could have in-principle been constructed from all passes through the reservoir. In summary, the work in Spagnolo et al. leverages channels in space to construct high-dimensional and correlated features, whereas we use time.

On operability and scalability, as we show in Fig. S18, the scheme we introduce in this work is greatly benefitted by an increased number of qubits to the system. Examples of such many-qubit-one-oscillator systems previously experimentally realized include Refs [1-3]. One could also extend our qubit-oscillator system to single-qubit-many oscillators [4,5] or many-qubits-many-oscillators [6], though these latter two configurations were not explored in simulations.

1. Song, C., Xu, K., Liu, W., Yang, C.-P., Zheng, S.-B., Deng, H., ... Pan, J.-W. (2017). 10-Qubit Entanglement and Parallel Logic Operations with a Superconducting Circuit. *Phys. Rev. Lett.*, *119*, 180511. doi:10.1103/PhysRevLett.119.180511
2. Paik, H., Mezzacapo, A., Sandberg, M., McClure, D. T., Abdo, B., Córcoles, A. D., ... Chow, J. M. (2016). Experimental Demonstration of a Resonator-Induced Phase Gate in a Multiqubit Circuit-QED System. *Phys. Rev. Lett.*, *117*, 250502. doi:10.1103/PhysRevLett.117.250502
3. Marinelli, B., Luo, J., Ren, H., Niedzielski, B. M., Kim, D. K., Das, R., ... Siddiqi, I. (2023). Dynamically Reconfigurable Photon Exchange in a Superconducting Quantum Processor. *arXiv [Quant-Ph]*.

4. Chakram, S., He, K., Dixit, A.V. et al. Multimode photon blockade. Nat. Phys. 18, 879–884 (2022)

5. Naik, R.K., Leung, N., Chakram, S. et al. Random access quantum information processors using multimode circuit quantum electrodynamics. Nat Commun 8, 1904 (2017).

6. Zhou, C., Lu, P., Praquin, M. et al. Realizing all-to-all couplings among detachable quantum modules using a microwave quantum state router. npj Quantum Inf 9, 54 (2023)

> **Action taken:** We have added the following sentences to the discussion: "Generalizing our approach to spatial in addition to temporal inputs, as was explored in Ref. [24], would likely support more sophisticated computations. In Appendix F, we explore such extensions in simulations and find a marked improvement in classification accuracy."

3) In my previous report, I asked the authors about the processing speed of the system, but I did not receive an answer. Since the system is an analog one, could the authors report the input rate in inputs per second or in analog samples per second?

We apologize for not answering the question clearly. We find it difficult to provide an answer that is expressed explicitly as samples per second, however, can we try to approximate it using the experimental results and the methods.

In Fig. 4 of the main text, we performed classification of noisy signals with different correlation lengths. We empirically showed that we could accurately distinguish between signals whose correlation lengths were on the order of 50 nanoseconds, or signals constructed with a minimum of 20 MSps sampling rate. This demonstrates that our quantum reservoir is sensitive to analog signals with changes on the order of 50 nanoseconds.

On the other hand, the time to perform the "processing", or classification takes a much longer time. The qubit is measured once every 1.2 microseconds, where each measurement gives us an output sample. These samples are then retrieved repeatedly to generate the statistics to form our output features. Once enough samples are obtained, we can accurately perform the classification among the different signals – for the tasks in this work, this is on the order of thousands of samples. Thus, while our analog reservoir computer is sensitive to signals that change on the order of 10s of nanoseconds, the speed at which we process the signals to perform the classification is much longer.

> **Action taken:** We have added the following sentences to the results section: "Additionally, the ability for our quantum reservoir to distinguish between signals with correlation times on the order of 50 ns demonstrates the sensitivity to signals which vary on time-scales much faster than the measurement rate."

REVIEWERS' COMMENTS

Reviewer #1 (Remarks to the Author):

My comment has been addressed in an excellent manner. I support the publication of the manuscript.

Reviewer #3 (Remarks to the Author):

The authors have properly addressed my previous concerns and I find the revised manuscript suitable for publication.