Supplementary Information for "Microwave signal processing using an analog quantum reservoir computer"

Alen Senanian[1,2], Sridhar Prabhu[1,2], Vladimir Kremenetski[2], Saswata Roy[1,2], Yingkang Cao[3,4], Jeremy Kline[2,‡], Tatsuhiro Onodera[2,5], Logan G. Wright[2,5,†], Xiaodi Wu[3,4], Valla Fatemi[2], Peter L. McMahon[2,6]

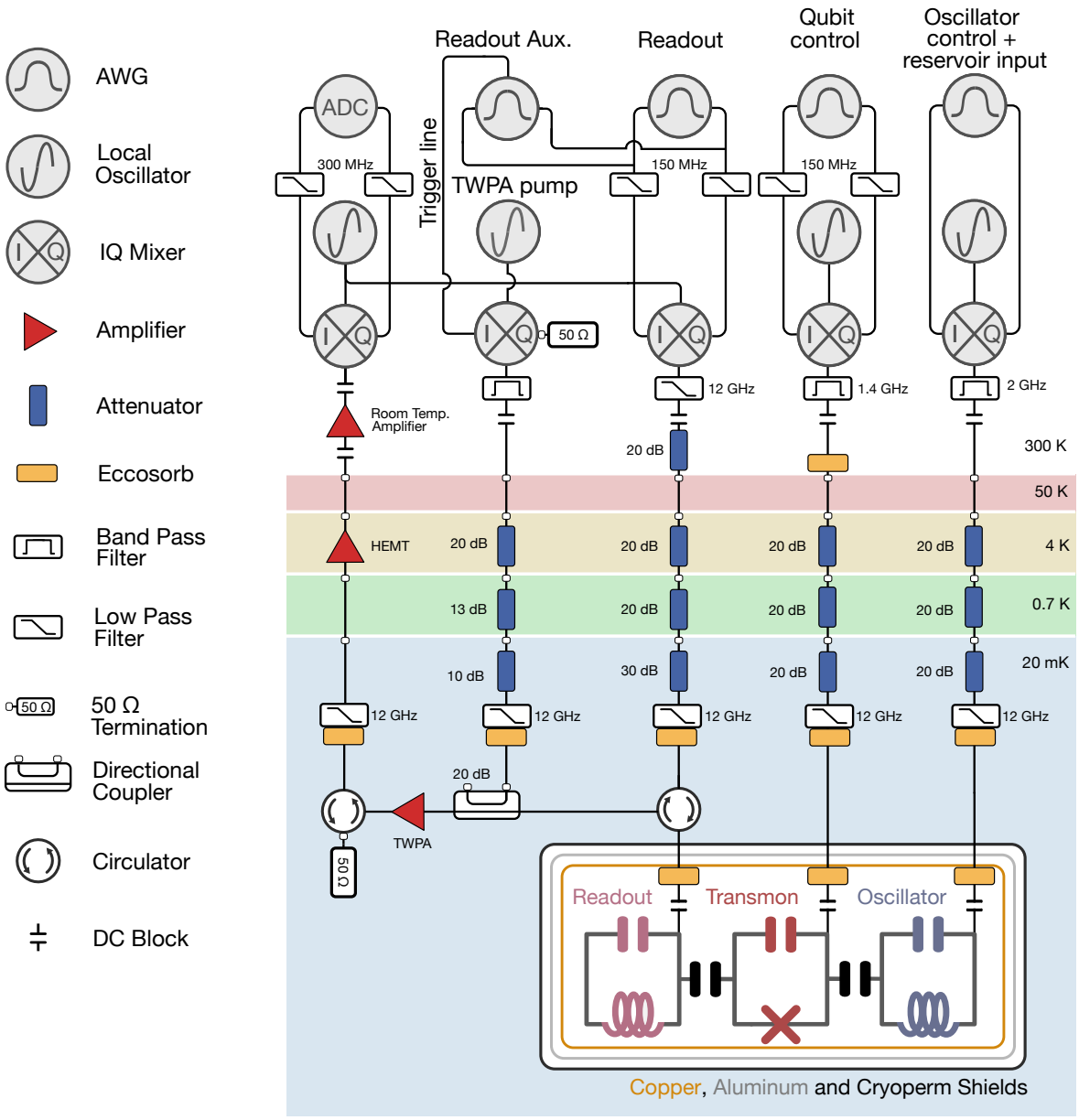
[1]Department of Physics, Cornell University, NY, USA.
[2]School of Applied and Engineering Physics, Cornell University, NY, USA.
[3]Department of Computer Science, University of Maryland, MD, USA.
[4]Joint Center for Quantum Information and Computer Science, University of Maryland, MD, USA.
[5]NTT Physics and Informatics Laboratories, NTT Research, Inc., CA, USA.
[6]Kavli Institute at Cornell for Nanoscale Science, Cornell University, NY, USA.

Contents

---

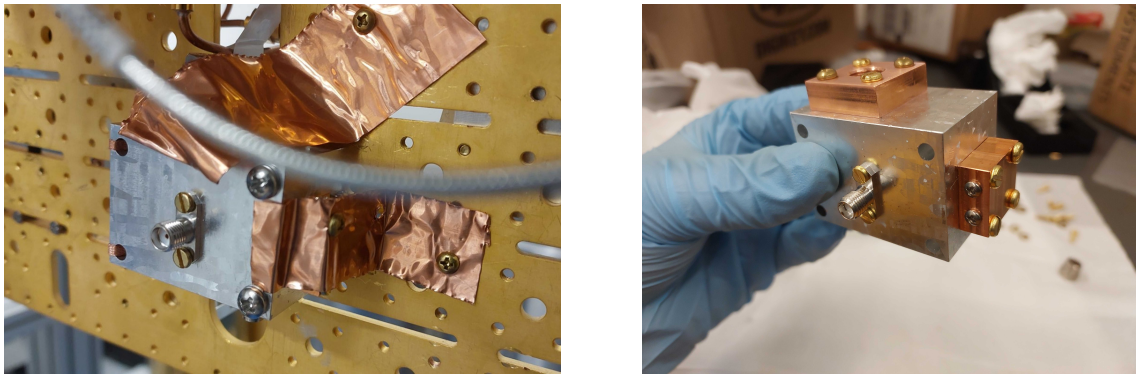
‡Present address: Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, MA, USA
†Present address: Department of Applied Physics, Yale University, CT, USA.

Supplementary Note 1 – Experimental setup



Supplementary Figure 1 Wiring diagram. Experimental setup for control hardware, cable routing, and shielding for our device.

The device used in this paper consists of an oscillator, a 3D stub post cavity made from high-purity 4N Aluminum treated with an acid etch, and a transmon qubit. The transmon, made of Niobium, is fabricated on a resistive silicon chip, along with an on-chip readout resonator also made of Niobium. The single chip hosting the transmon and the readout resonator is mounted in the 3D cavity package using copper clamps. The cavity and the copper clamp contain copper films for thermalization directly to gold-plated copper breadboard at the mixing chamber plate of the dilution refrigerator (Supplementary Figure 2). The device is shielded with Copper coated with Berkeley Black, and two types of magnetic shields: Aluminum, and Cryoperm (Supplementary Figure 1). The cavity pin is set such that the oscillator mode is undercoupled to the transmission line by a factor of 40. While this reduces the



Supplementary Figure 2 Photo of device. The device consists of a on-chip transmon and a co-planar waveguide readout resonator mounted inside a high-purity Aluminum cavity. The package is mounted to a gold-plated copper breadboard at the mixing-chamber plate of a dilution refrigerator.

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.

The control pulses for the qubit and the storage are synthesized using Zurich Instruments (ZI) HDAWG, which have a baseband bandwidth of 1 GHz. These are upconverted using Rohde & Schwarz SGS100A, which are signal generators with built-in IQ mixers. These built-in mixers are used for all frequency conversions with the exception of the readout. The readout pulses are synthesized and digitized using ZI UHFQA, and are up-converted and down-converted using Marki mixers (MMIQ-0416LSM-2), with a split LO from a single SGS100A. Readout signals are first amplified with a traveling-wave Josephson Amplifier (TWPA), which is a quantum-limited amplifier. The TWPA typically requires large pump tones, so we gate it with a trigger line from the readout AWG which combines with the CW pump tone in an IQ mixer (as a makeshift fast switch). The readout signals are then further amplified with a High-electron mobility transistor (HEMT) ampliflier at the 4K stage, and again amplified with a room temperature amplifier (ZVA-1W-103+ from Mini-Circuits) and filtered. The digitizer on the ZI UHFQA converts to the analog response to a digital signal and integrates it to produce a binary outcome depending on the qubit state.

For the experiments that intentionally suppress the qubit T_2 via resonator induced dephasing via pumping of the readout resonator, we use an additional ZI HDAWG channel that combines with the AWG of the ZI UHFQUA. This was mostly a choice out of convenience, as the AWG of the ZI UHFQUA has limitations that made characterizations tricky.

Supplementary Note 2 – System Hamiltonian & Reservoir description

Hamiltonian description

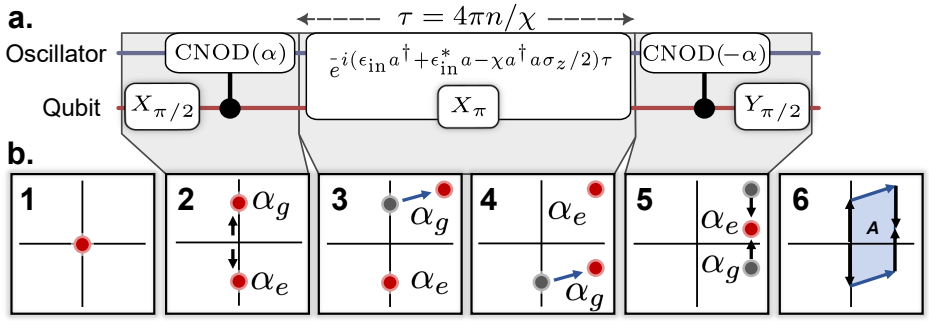
Our transmon-cavity system is well approximated by the Hamiltonian [1]:

$$H/\hbar = \omega_q q^\dagger q + \omega_a a^\dagger a - \chi q^\dagger q a^\dagger a - \chi' q^\dagger q a^{\dagger 2} a^2 - K_q q^{\dagger 2} q^2 - K a^{\dagger 2} a^2 + \Xi(t)(q + q^\dagger) + \xi(t)(a + a^\dagger), \tag{1}$$

where a is the annihilation operator for the oscillator mode, and q is the annihilation operator for the qubit mode, ω_a and ω_q are the frequencies of the oscillator and qubit mode respectively, χ and χ' are the dispersive shift and the oscillator state-dependent dispersive shift respectively, K and K_q are the self-Kerr of the oscillator and the transmon anharmonicity respectively. The values for these parameters, as well as values for decay rates, are listed in Supplementary Table 1. 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 of the main text via $\Xi(t) = \Omega(t)e^{i\omega_q t} + \text{H.c.}$ and $\xi(t) = \epsilon(t)e^{i\omega_a t} + \text{H.c.}$. For the design of our drives, we ignore the self-Kerr of the oscillator as well as the higher-order cross-Kerr. We note that these are indeed present, but for the purposes of a quantum reservoir, only add to the complexity of the dynamics. Finally, moving to the rotating frame of the transmon and oscillator mode and truncating to the first two levels of the transmon, we arrive at the Hamiltonian in Eq. 1 of the main text.

Parameter	Mode(s)	Symbol	Value
Frequency	Transmon g-e	ω_q	$2\pi\times$ 7.136 GHz
	Oscillator	ω_a	$2\pi\times$ 6.024 GHz
	Readout	ω_r	$2\pi\times$ 8.888 GHz
Self-Kerr	Transmon g-e	K_q	$2\pi\times$ 315 MHz
	Oscillator	K	$2\pi\times$ 6 kHz
Cross-Kerr	Transmon-Oscillator	χ	$2\pi\times$ 2.415 MHz
	Transmon-Readout	χ_{qr}	$2\pi\times$ 1 MHz
Second-order Cross-Kerr	Transmon-Oscillator	χ'	$2\pi\times$ 19 kHz
Relaxation time	Transmon g-e	T_1	30 μs
	Oscillator	T_1^a	100 μs
Dephasing time	Transmon g-e	T_2	25 μs
Thermal population	Transmon g-e	$\bar{n}_{\rm eq}^q$	3%
	Oscillator	$\bar{n}_{\rm eq}^a$	$< 0.2\%$

Supplementary Table 1 System parameters and dissipation rates. System parameters were measured using various spectroscopic and time-domain techniques following methods in Ref. [2].



Supplementary Figure 3 Geometric phase unitary to sense phase of unknown displacement (a.) Decomposition of the unitary used throughout the reservoir for time-independent tasks. **(b.)** Schematic representation of the dynamics of the oscillator state under the reservoir drives with time-independent input, highlighting a unitary which implements a geometric phase unitary. (1) At the start of the protocol, the qubit is in the ground state, with the oscillator at vacuum. (2) An initial $X_{\pi/2}$ pulse brings the qubit to the equator. The CNOD(α) unitary conditions the state of the oscillator based on the qubit. For the first reservoir, this is a coherent state. (3) and (4) For time independent inputs, the effective action can be described by a single displacement on the oscillator mode. In this experiment, we operate the displacement at a frequency which, to first order, causes a displacement only on the state conditioned on the ground state of the qubit. A qubit π pulses switches the state in between the two displacements. (5) The final conditional displacement brings the two conditioned states back onto each other. This effectively disentangles the qubit from the oscillator mode. (6) The effective geometrical area enclosed A, which is a function of the input, is imparted onto the qubit.

Reservoir description for time-independent signals

The advantage of the reservoir computing paradigm is the flexibility in the choice of dynamics. However, simple design principles, motivated by the physics of the system, can go a long way in engineering a reservoir with high expressive capacity on many tasks. In this section, we provide full details and motivations for the unitaries and measurements in this work, followed by sections outlining characterizations of the device in order to realize the intended dynamics.

The reservoir drives consists of two categories of dynamics: the unitaries and the measurements. In what follows, we will first provide analysis of the dynamics for time-independent input (e.g. the signals in Fig. 2). As we will see, the unitary component of the dynamics implemented in this work strives to implement a $\cos^2$ nonlinearity on the raw input, whereas the measurements generate non-classical features in the state and quantum correlations in the measurement trajectories via measurement backaction.

While measuring the quadratures of some unknown signal is easy with a typical homodyne setup, performing the same measurement of a displacement on an oscillator using only qubit measurements can be non-trivial. Of course, when designing a reservoir, one does not strive to implement the identity, but it is a good starting point – the unitary is thus implemented to approximate the identity. It consists of the input signal data, which is sandwiched on

either side by fast conditional displacement gates implemented with CNOD [3] and qubit rotation gates. The broad-overview of the decomposed unitary is given in terms of gates in Supplementary Figure 3, along with a schematic portrayal of the phase-space trajectory of the oscillator mode initialized in vacuum subject to a time-independent drive.

We begin with an idealized gate-based version decomposition of our reservoir for time-independent input on resonance with the oscillator conditioned on the qubit being in the ground state. The sequence of gates the reservoir unitary approximates:

$$U_1 = X_{\pi/2} \tag{2}$$
$$U_2 = D(\alpha)|g\rangle\langle e| + D(-\alpha)|e\rangle\langle g| \quad \text{CNOD} \tag{3}$$
$$U_3 = D(\beta)|g\rangle\langle g| + |e\rangle\langle e| \quad \text{Input} \tag{4}$$
$$U_4 = X_{\pi} \tag{5}$$
$$U_5 = U_3 = D(\beta)|g\rangle\langle g| + |e\rangle\langle e| \quad \text{Input} \tag{6}$$
$$U_6 = D(-\alpha)|g\rangle\langle e| + D(\alpha)|e\rangle\langle g| \quad \text{CNOD} \tag{7}$$
$$U_7 = Y_{\pi/2} \tag{8}$$

Ignoring the very first unitary, after applying the sequence of unitaries U_2 through U_7, we arrive at unitary

$$U_7U_6U_5U_4U_3U_2 = \frac{i}{\sqrt{2}}e^{(\alpha\beta^* - \alpha^*\beta)}D(\beta)(|g\rangle\langle g| - |e\rangle\langle g|) - \frac{i}{\sqrt{2}}e^{(-\alpha\beta^* + \alpha^*\beta)}D(\beta)(|g\rangle\langle e| + |e\rangle\langle e|) \tag{9}$$

Let

$$|\psi\rangle = [e^{-i\phi/2}\cos(\theta/2)|g\rangle + e^{i\phi/2}\sin(\theta/2)|e\rangle] \otimes |\text{cavity}\rangle \tag{10}$$

be some arbitrary initialized state. Then for $\theta = \pi/2$, we have

$$U_7U_6U_5U_4U_3U_2|\psi\rangle = \frac{1}{\sqrt{2}}D(\beta)[i\sin(A - \phi/2)|g\rangle + \cos(A - \phi/2)|e\rangle] \otimes |\text{cavity}\rangle, \tag{11}$$

where $A = 2|\alpha||\beta|\sin(\delta) = i(\alpha\beta^* - \alpha^*\beta)$ is the geometric phase enclosed by the oscillator trajectory which is dependent on the phase difference δ between the known displacement $D(\alpha)$, and the unknown displacement $D(\beta)$ (Fig 3b). Thus, for the proper qubit state before the application of $U_2 \dots U_7$, we are able to extract information about the phase of the displacement. We also note that the qubit and the oscillator are disentangled after the unitary, and that the effect of the unitary on the oscillator mode is a simple displacement.

Finally, pre-pending U_1 (Eq. 2) to the string of unitaries guarantees that we initialize our qubit state with $\theta = \pi/2$ when following a qubit measurement, independent of that measurement outcome. It also guarantees $\phi = \pi/2$ or $3\pi/2$ depending on the measurement outcome. The probability of measuring the qubit in the excited state conditioned on preparing it e vs g after the entire sequence is then:

$$P_{e|g} = \cos(A - \pi/4)^2 \qquad P_{e|e} = \sin(A - \pi/4)^2 \tag{12}$$

Thus, with this sequence of unitaries, we are able to extract the phase of some unknown displacement (relative to some known displacement α) by simply measuring the qubit. While for the first run of the reservoir, the qubit will start in the ground state (up to thermal noise), after performing a parity measurement, the qubit state will depend on the previous measurement outcome. See Supplementary Figure 7 for an experimental implementation of the above results.

In principle, Eq. 12 enables us to perform the identity operation on the input x, y points followed by a $\cos^2$ kernel. Without loss of generality, we take $\arg(\alpha) = 0$, then $i(\alpha\beta^* - \alpha^*\beta) = \text{Im}(\beta) = \beta_x$. Alternating between $\arg(\alpha) = 0$ and $\arg(\alpha) = \pi/2$ allows us to extract $\cos^2(\vec{\beta})$ with two runs of the reservoir.

Whereas all gates besides the input (Eqs. 4 and 6) are fast and therefore insensitive to the cross-Kerr interaction, the primary deviation from the gate description occurs for the input, which can be very long. This input displacement is conditioned on the qubit being in the ground state. Therefore, in the rotating from of the qubit-oscillator system, the branch of the oscillator state conditioned on the qubit being in the excited state will rotate at a frequency χ, which in general will break the geometric phase construction that works for time-independent tasks. Therefore, we

limit the exposure time of the reservoir to the input signal to be an integer multiple of $4\pi/\chi$, so that the oscillator state conditioned on the qubit being in the excited state will return to the same point.

The unitary described in Eqs. 2-8 is followed by a qubit measurement, then a parity measurement Π [4, 5] with projectors $P_\pm$, where

$$\Pi = (-1)^{a^\dagger a} \qquad P_\pm = \frac{1}{2}(1 \pm \Pi) \tag{13}$$

As mentioned above, the effect of the unitary on the oscillator state for time-independent signals is simply a displacement of the input data $D(\beta)$, independent of the qubit measurement outcome. For the following discussion, we will ignore the qubit dynamics, since the qubit and the oscillator are disentangled at the end of the unitary. In effect, the state of the oscillator can be described by a series of alternating displacements and parity measurements:

$$|\text{cavity}\rangle = \ldots P_{p_4} D(\beta) P_{p_3} D(\beta) P_{p_2} D(\beta) P_{p_1} D(\beta) |0\rangle, \tag{14}$$

where P_{p_n} is the projector of the nth parity measurement with outcomes $p_n = \{+, -\}$. For k runs of the reservoir, we can reorder terms and add pairs of canceling displacements $D(-\beta)D(\beta)$ to rewrite the above as

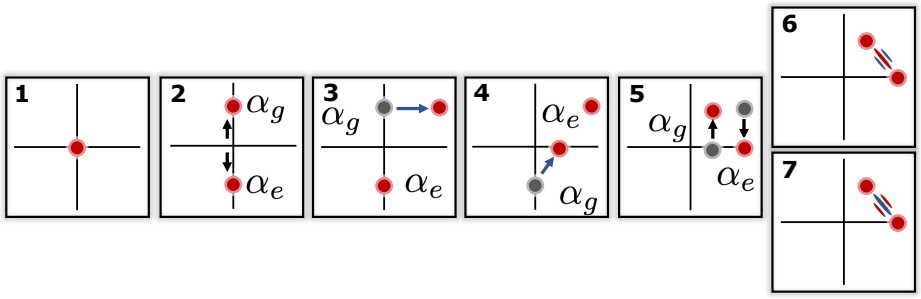
$$|\text{cavity}\rangle = \left(\prod_n^k P_{p_n}^{n\beta} \right) D(k\beta)|0\rangle. \tag{15}$$

Equation 15 describes a series of projective measurements after preparing a displaced vaccum state. The projectors and their associated measurements are

$$P_\pm^\alpha = D(\alpha) P_\pm D(-\alpha) \qquad \Pi^\alpha = D(\alpha) \Pi D(-\alpha) \tag{16}$$

The measurements Π^α describe parity measurements in displaced frame at α. Incidentally, the expectation value of this operator are proportional to the Wigner function at α [6]. However, importantly, Eq. 15 does not describe performing Wigner tomography of the state $D(k\beta)|0\rangle = |k\beta\rangle$ at points given by $\beta, 2\beta, 3\beta, \ldots$, as the effective measurements Π^α do not commute for different values of α. Instead, in general $[\Pi^\alpha, \Pi^\gamma] \neq 0$. Therefore, in this light, our reservoir construction can be seen to leverage non-commuting measurements and quantum contexuality to generate conditional and correlated probabilities over measurement trajectories.

Reservoir description for slow varying time-dependent signals



Supplementary Figure 4 Schematic description of the dynamics of the oscillator, starting in vacuum, for a slow time-varying input In such a scenario (as is the case for the task of classifying radio modulation schemes), the signal causes a displacement largely conditioned on the ground state. Generally, the value of the displacement is different before and after the π pulse in between (3) and (4). Unlike the regime for time-independent signals, there is no effective area enclosed in phase space, which leaves the qubit entangled with the oscillator. (6) and (7) describe the state of the oscillator after the qubit measurement. The resulting state of the oscillator is a cat state, where the parity of the cat dependent on whether the outcome of the qubit measurement is ground or excited.

For generic, time-dependent signals, like those classified in Figs. 3 and 4 in the main text, the geometric unitary described by Eqs. 2-8 does not in general hold, as the symmetry between panels 3 and 4 in Supplementary Figure 3

is broken. Additionally, the approximation that the input is displacement conditioned on the qubit in the ground state (Eq. 4 and 6) will not hold for high bandwidth signals, like those in Fig. 4 in the main text. For high-bandwidth signals, the input will also have some contribution in displacing oscillator conditioned on the qubit being in the excited state, which can lead to complex dynamics in the oscillator. While for generic signals, this can be hard to describe, here we prove a treatment of our reservoir construction for slowly-varying, time-dependent signals, like those in Fig. 3 of the main text.

We can follow most of the derivation from the scenario of time-independent signals, to describe the dynamics of the QRC for the task of classifying radio frequency modulation schemes. Along with the assumptions in the previous section, we make the slow-varying input approximation, such that the displacement on the oscillator of the reservoir is still effectively conditioned on the ground state of the qubit. The displacement on the oscillator depends on the value of the symbol encoded for the given modulation scheme. Since, in general, the symbol is different before and after the qubit π pulse: the direction of the displacement in the oscillator will be different. Given the timescales of the input signal involved, this essentially corresponds to a displacement on the oscillator conditioned on the ground state of the qubit. When the two displacements are different in magnitude and direction, the qubit remains entangled with the oscillator at the end of the reservoir unitary. The state of the system just before the measurements is (step (5) of Fig 4:

$$|\psi\rangle = \frac{1}{\sqrt{2}}(e^{-iA_i}D(\beta_i)|g, \text{cavity}\rangle + e^{iA_j}D(\beta_j)|e, \text{cavity}\rangle), \tag{17}$$

where β_i is the displacement before the qubit flip, and β_j is the displacement after. $A_i = 2\text{Im}\left[\alpha\beta_i^*\right]$ is the phase acquired after two non-orthogonal displacements. When $\beta_i = \beta_j$, we recover the dynamics for time independent signals. It is straightforward to show that the qubit will be disentangled from the oscillator and that the area A_i, corresponding to the geometrical phase form the area enclosed in phase space will be present as a relative phase difference between the ground and excited state. After a $Y_{\pi/2}$ gate, we have the following state in our system:

$$|\psi\rangle = \frac{1}{2}[e^{iA_i}D(\beta_i) + e^{-iA_j}D(\beta_j)]|g, \text{cavity}\rangle + \frac{1}{2}[e^{iA_i}D(\beta_i) - e^{-iA_j}D(\beta_j)]|e, \text{cavity}\rangle, \tag{18}$$

One can think of this as a cat state in the cavity, with a parity determined by the qubit state. This is schematically shown in (6) and (7) in Fig 4. In the limit of very different displacements, the probability of the qubit measurement is the same for both ground and excited states. The goal of this task can be thought of as discriminating probability distribution functions over the (I, Q) plane. Fig 3a represents the so-called "constellation " diagram of the modulation schemes considered in this work. Each scheme can take discrete values in (I, Q) space, with even equal probability (we construct the dataset of radio signals encoding random binary strings). Our lack of knowledge of the exact displacement on the oscillator can be mathematically expressed as a density matrix. This is the most apparent in the state of the oscillator after the initial qubit measurement,

$$\rho'_{\text{cavity}} = \sum_{\beta_i \in P} p_i D(\beta_i)\rho^\dagger_{\text{cavity}}D(\beta_i), \tag{19}$$

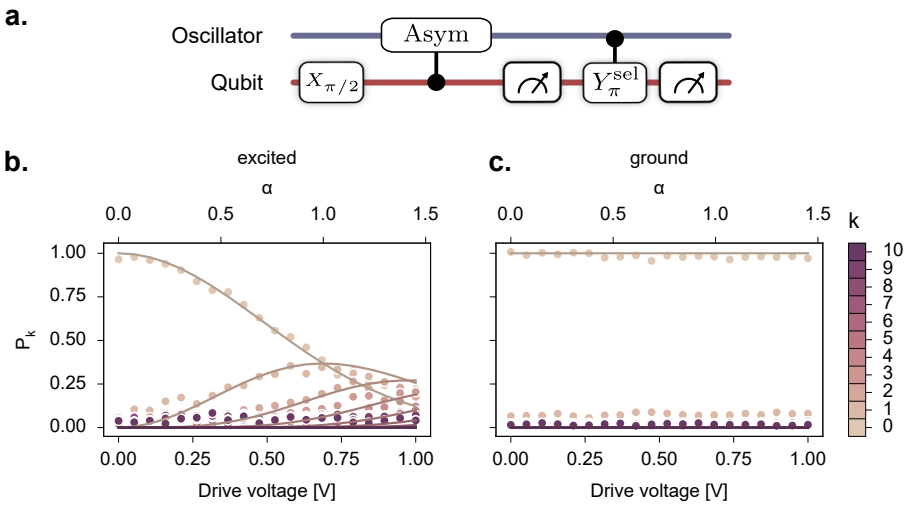
where ρ'_{cavity} is the density matrix representation of the cavity right after the qubit measurement and ρ_{cavity} describes the initial density matrix before the application of the protocol. The set P describes the distribution of possible displacements which can be received from the input. p_i is probability for receiving the symbol corresponding to a displacement β_i For the task considered in this work, these distributions are uniform, with no contributions from conditional probabilities. However, this description of the reservoir motivates the potential for the QRC to distinguish signals with complex correlations in the symbols of the message encoded.

Supplementary Note 3 – Quantum reservoir characterization

CNOD

Here, we provide the calibration of the CNOD unitary [3], one of the components of our reservoir unitary (Supplementary Figure 3). The CNOD protocol implements the following unitary

$$\text{CNOD}(\alpha) = D(\alpha)|g\rangle\langle e| + D(-\alpha)|e\rangle\langle g|. \tag{20}$$



Supplementary Figure 5 Characterization of conditional displacements using an Anti-symmetric pulse from Ref. [3]. (a.) Pulse control schematic of characterization of conditional displacement. Here, we make use of number-splitting spectroscopy to characterize the state of the cavity after perform a conditional displacement on the state $|+\rangle|\text{vacuum}\rangle$ conditioned on the qubit state. By post-selecting on the qubit state, we can evaluate the effectiveness of the conditional displacement. **(b.)** Number splitting spectroscopy conditioned on measuring the qubit in the excited state. A single parameter fit is used to capture the behavior of the state of the cavity as a function of amplitude. From the good agreement, we conclude that conditioned on measuring the qubit in the excited state, the cavity is displaced. **(c.)** Number splitting spectroscopy conditioned on measuring the qubit in the ground state. We see very limited change in the cavity state when measuring the qubit in the ground state.

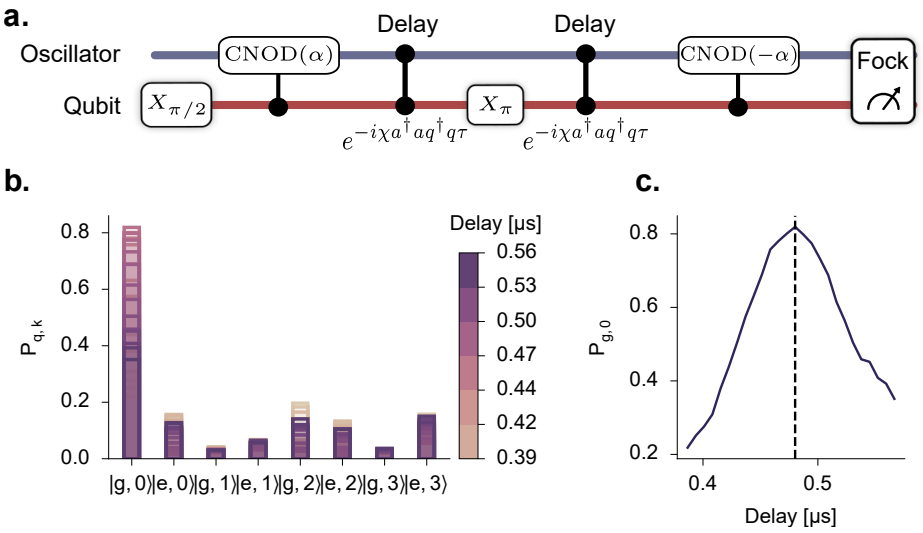
The protocol is implemented with two 'Anti-symmetric pulses' sandwiching a qubit pi-pulse. In the frequency domain, the pulse is composed of two gaussian envelopes offset such that there is a zero-crossing at the qubit ground state frequency, and that the spectrum is anti-symmetric around this point (see Ref. [3]). The Anti-symmetric pulse is a conditional displacement, conditioned on the qubit being in the excited state. The motivation for using CNOD instead of a single tone displacement on resonance with the stark-shifted qubit frequency is that it enables the ability to perform conditional displacements at time scales much smaller than $2\pi/\chi$.

Supplementary Figure 5a displays the protocol for characterizing the anti-symmetric pulse. First, the qubit is unconditionally brought to the equator of the bloch sphere, with a wide-band $X_{\pi/2}$ pulse. After this, the anti-symmetric pulse acts on the cavity, followed by an qubit measurement, collapsing the cavity state to either $D(\alpha)|0\rangle$ or $|0\rangle$. After collapsing the state, we perform a number-splitting spectroscopy on the cavity. This is performed with a conditional Y_{π}, conditioned on the kth cavity Fock state [7, 8] followed by a second qubit measurement. By post-selecting on the first qubit measurement outcome, we can characterize the cavity state for each branch. Supplementary Figure 5b and c show the number-splitting spectroscopy for the cavity state conditioned on the qubit being in the ground state vs excited, as a function of pulse amplitude. These curves are fitted with a single parameter scaling parameter that defines the relationship between pulse amplitude voltage and the amount of displacement α.

Reservoir unitary characterization

With our rotation gates and CNOD's calibrated, we describe in this section the calibration of signal drives toward the implementation of Eqs. 2-8. We begin with a calibrating the duration of time our reservoir is exposed to the input signal. Calibrating this delay is crucial for a faithful implementation of the geometric phase detection unitary introduced in this work. While it may seem that this restriction in the input signal duration is contrived in a real-world setting where the signal is unknown, this restriction would be implemented via a fast switch that exposes our device to the unknown signal periodicially.

Supplementary Figure 6a schematically describes the experimental protocol for calculating the delay between the two CNOD pulses. Here, we effectively try to undo a double conditional displacement via second double conditional displacement. Due to the dispersive shift, after the first conditional displacement, the state of the cavity conditioned



Supplementary Figure 6 Characterization of input signal duration **(a.)** Diagram of control sequence to calibrate length of input signal duration toward implementing the geometric phase unitary in Eqs. 2-8. Here, a double conditional displacement is performed after sending the qubit to the equator of the Bloch sphere. After this, a variable delay is added before undoing that displacement. Finally, the qubit and the Fock distribution of the cavity is sampled using methods from Refs. [9, 10]. **(b.)** Overlapping histograms showing Fock distribution of cavity state conditioned on qubit state as a function of delay. **(c.)** Cavity state overlap with the vacuum state as a function of delay. At a particular value of the delay, the two displacements interfere and cancel each other out.

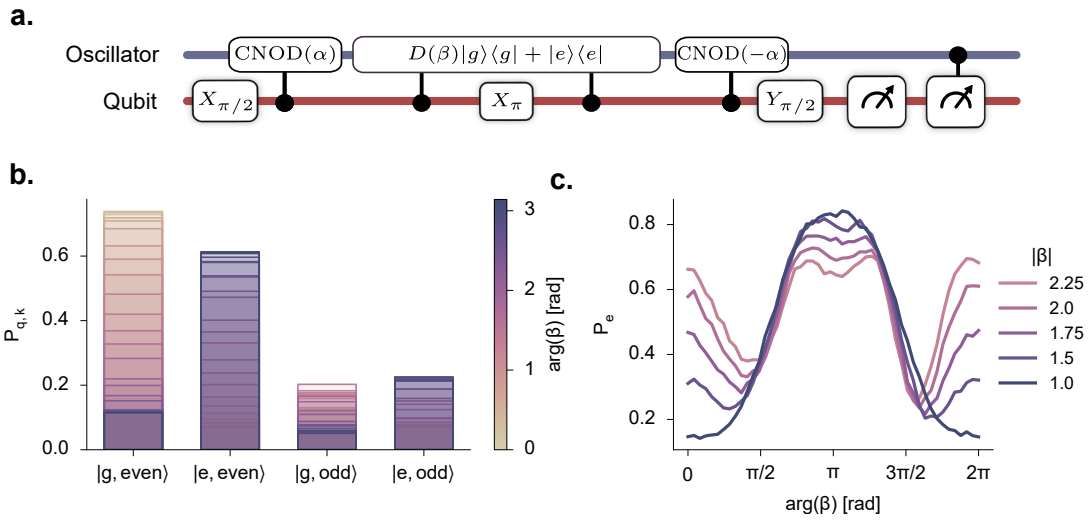
on the excited state of the qubit will start rotating with respect to the state of the cavity conditioned on the ground state. After a period of $2\pi/\chi$, the this will return to the same position as the start. Undoing the displacement at this point in time will send the cavity state to vacuum. Supplementary Figure 6b shows the Fock distribution of the cavity as a function of the waiting time, and Supplementary Figure 6c shows the cavity state overlap with the vacuum state as a function of the waiting time.

Next, we implement the full unitary given by Eqs. 2-8, where the section corresponding to the input data displacement (Eqs. 4 and 6) is given by the duration found in the results above. For this calibration, we implement the full unitary given by the diagram in Supplementary Figure 3a by varying the angle of the input displacement and looking at the dependence.

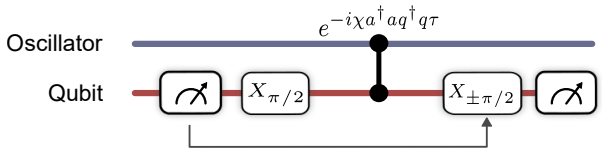
Supplementary Figure 7a shows the schematic overview of the calibration procedure. The geometric phase unitary is parameterized by a long displacement, whose angle we sweep. After the unitary we perform a qubit measurement, followed by a parity measurement. This calibration experiment is essentially identical to the time-independent reservoir computing experiments in terms of the control protocol. Here, instead of sending data from different distributions for the system to classify, we only vary the phase and amplitude of some input displacement to get the phase dependence we want.

Supplementary Figure 7b shows the distribution of measurement outcomes from measuring the qubit and the oscillator parity after the unitary is applied with $\alpha = 1$ and $\beta = 0.25$. As the angle of the input is swept, the qubit probability of the qubit being in found in the ground state shifts to being found in the excited state. This is more evident in Fig.7c where we plot the probability of measuring the qubit in the excited state P_e as a function of the phase of β for different amplitudes of β. In comparison we find great qualitative agreement with the expected result $P_e = \cos(2|\alpha||\beta|\cos(\delta) + \pi/4)^2$, where $\delta = \arg(\alpha) - \arg(\beta)$ (see Eq.12), though we find an extra reduction in the dynamic range in P_e for increasing β due to qubit overheating.

For our quantum reservoir tasks, we choose α to be quite small, near 0.2. The effect of this is a severe reduction in the dynamic range of P_e, but one that is easily distinguishable at 1000 shots. For all of our tasks, this was the mininum number of shots needed to get 100%. Keeping $|\alpha|$ small allows for a greater sensitivity in $|\beta|$ without worrying about qubit overheating.



Supplementary Figure 7 Full geometric phase unitary calibration. **(a.)** Pulse control protocol for the unitary defined by Eqs 2-8. This protocol is identical to that of Supplementary Figure 3a. **(b.)** Overlapping histograms showing the probability over outcomes of measuring the qubit and oscillator-parity as a function of the phase of the input signal with $|\beta| = 0.25$. As the phase is varied, the probability of the qubit being excited increases. **(c.)** The probability of measuring the qubit excited as the phase of the input displacement β is varied, plotted for different values of $|\beta|$



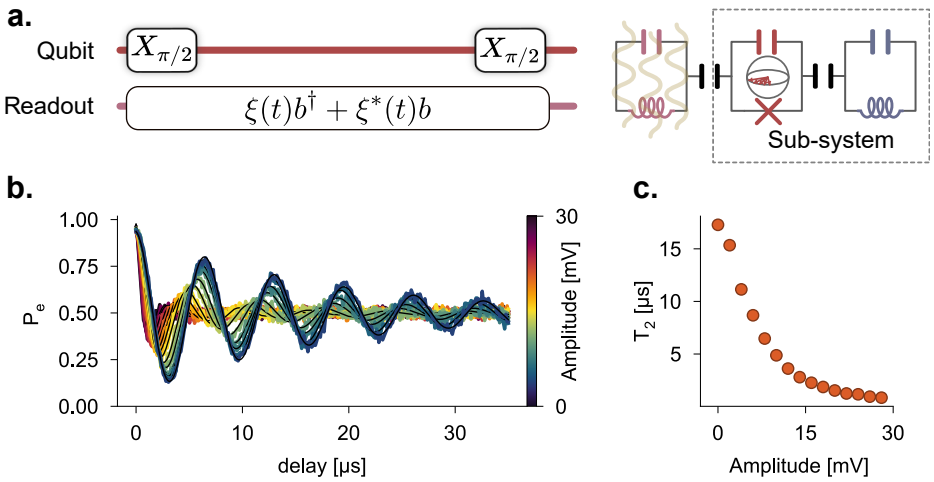
Supplementary Figure 8 Qubit and oscillator-parity measurements in the quantum reservoir computer. For our reservoir construction, the state of the qubit is not generally known before a parity measurement. We apply feedback to change the parity measurement based on the preceding qubit measurement to faithfully capture the oscillator parity.

Qubit & parity measurements

The qubit and parity measurements performed in this work are the standard pulse schemes used in many previous works, with one change. The typical procedure of measuring the parity of a cavity state is similar to a Ramsey experiment (and perhaps more closer still to a 'qubit-revival' experiment [2]), and importantly requires knowledge of the state of the qubit before the measurement is performed. In a quantum reservoir setting where measurement trajectories can be unknown, measuring the parity of the cavity is not straight-forward without post-selection or feedback. Here, since we perform a qubit measurement just before the parity measurement, we apply simple feedback that conditions the parity unitary on the measurement outcome of the preceding qubit measurement. The condition is such that the parity measurement outcome is now independent of the preceding measurement outcome. This reduces the order of correlations required to gain the same information: attaining the parity of the cavity only requires information about the parity measurement, whereas previously, second-order correlations between the qubit and parity measurement was required. A further refinement to reduce trivial correlations in the measurement history would reset the qubit after the oscillator parity, however, due to limitations in the FPGA software, this was not implemented.

Tuning T_2 via resonator-induced dephasing

Here we describe the experiment to reduce the qubit coherence time by pumping the readout resonator with photons during our reservoir experiments (see Fig. 2d). The calibration of this experiment involves performing a standard Ramsey T_2 experiment, modified with a pump on the readout resonator (Supplementary Figure 9a). Once populated,



Supplementary Figure 9 Resonator-induced dephasing via pumping on the readout resonator (a.) Protocol for characterizing the effect of a readout pump on qubit decoherence: we perform a typically Ramsey experiment while populating the readout resonator to measure the T_2. **(b.)** Ramsey curves as a function of readout pump amplitude. These curves are fit using Eq. 21 to produce estimates of the qubit coherence. **(c.)** Extracted T_2 values for each of the curves in part (b)

the resonator photons induce a dispersive shift, which sends the qubit to the center of the Bloch sphere once the readout resonator is traced out. In principle, this interaction is coherent, and the qubit should see a revival. However, due to the leaky nature of the readout resonator by design, a coherent revival is not observed. As remarked at the end of Supplementary Note 1, this experiment required an auxiliary AWG line. Supplementary Figure 1 denotes this as the 'Readout Auxiliary' line.

Supplementary Figure 9b shows the results of the Ramsey calibration with the readout pump on, for varying pump powers. We see a steady decrease in the qubit coherence time as the pump amplitude is increased as expected. The curves are fit to the equation

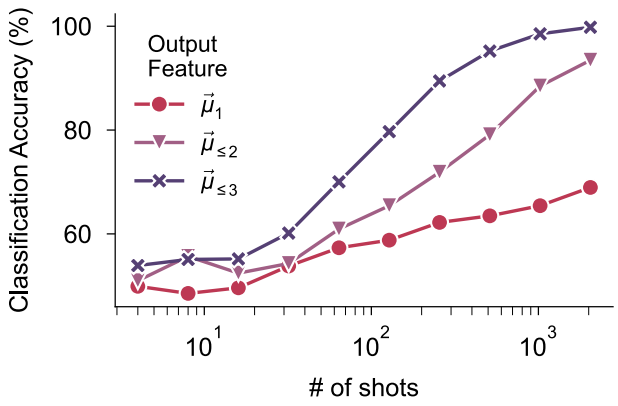
$$P_e = \cos(2\pi\delta t)e^{-t/T_2}, \tag{21}$$

where δ is an intentional detuning. Here, a Gaussian pulse was used as the readout pump. We expect that due to the construction of the reservoir, a flattop pulse may be more detrimental to the classification performance, since the Gaussian pulse has little amplitude during the CNOD unitaries shown in Supplementary Figure 3a. Finally, we note that the maximum T_2 shown in Supplementary Figure 9 differs from the value quoted in Supplementary Table 1. After preliminary calibration data corresponding to those in Supplementary Figure 9, the experiments in Fig. 2c were performed, after which the qubit T_2 was suddenly lowered. However, all experiments presented in this manuscript, with the exception of Supplementary Figure 9, were performed where the qubit T_2 matched that of Supplementary Table 1. Given the conclusion that the qubit T_2 does not impact classification accuracies until it approaches the time between measurements, we decided to include the higher quality data presented in Supplementary Figure 9, rather than the preliminary data used to calibrate the results in Fig. 2c.

Supplementary Note 4 – Machine learning with the quantum reservoir

Output feature encoding

In this work, we use measurement correlations as the output feature vectors from which the trained linear layer of our reservoir performs the classification. In this section, we provide details in how these were constructed from measurement results, as well as motivations and comparisons with other output encodings. As described in the main text, measurements of our reservoir involve two measurements following every data input: a qubit measurement and a parity measurement. The qubit measurement, which follows just after the input unitary, either extracts information about the input displacement (if the signal is time-independent), or performs some nontrivial back-action on the oscillator state (see Supplementary Figure 4). The parity measurement, which follows the qubit measurement, will simply measure the parity of the cavity state post-qubit measurement, and collapse the oscillator state to either



Supplementary Figure 10 Comparison of feature vectors in spiral classification performance Here, classification accuracy on the spiral task is considered for different output encodings. Particularly, we compare included higher and higher correlations. $\vec{\mu}_{\leq p}$ describes a feature vector containing all central moments up to and including the p-th central moment (see text).

even or odd Fock states. It is worth pointing out that measurements of the parity are done with an entangling unitary starting with a known qubit state and then performing a regular qubit measurement (see Supplementary Figure 8 for details).

In this manuscript, qubit measurements are performed using standard dispersive readout, which we review here, since the process involves a number of nonlinear steps (for a thorough review, see Ref. [1]). Each measurement outcome is the result of integrating a response signal from the readout resonator, and is defined by a single point on the $I-Q$ plane. For sufficiently strong coupling between the readout resonator and the qubit compared with the resonator linewidth, the set of all possible integrated IQ points will form two (or more) localized and well-seperated blobs, indicating projective measurement with single-shot fidelity. These two (or more) blobs correspond to different states of the transmon, and single-shot fidelity refers to the ability to discern the state of the qubit using only one readout pulse. With knowledge of the location of these blobs, and which state they correspond to, we perform a threshold the measurement result to either '0' or '1', indicating the qubit ground state or excited state respectively.

From a string of binary measurement outcomes, or bitstring, we form our feature vectors by first calculating the p-th central moment μ_p, defined as

$$(\mu_p)_{ijkl\ldots} = \frac{1}{N_{\text{shots}}} \sum_{n}^{\text{N}_{\text{shots}}} (x_{ni} - \langle x_i \rangle)(x_{nj} - \langle x_j \rangle)(x_{nk} - \langle x_k \rangle)(x_{nl} - \langle x_l \rangle)\ldots, \tag{22}$$

where the number of indices of μ_p is equal to p. Here x_{ni} is the nth repeated measurement result of observable x_i. In our setting, i labels the i-th measurement in a sequence of correlated measurements before the system is reset. The expectation value $\langle x_i \rangle$ is taken over the shots N_{shots} – counting the number of system resets and repetitions. Faithful estimates of these moments typically require on the order of 1000 shots for the results presented in this manuscript.

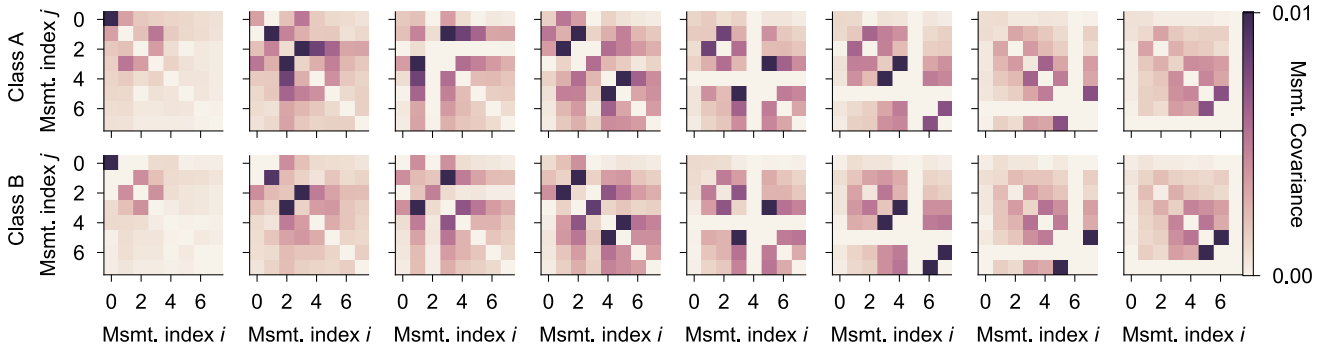
The central moments of Eq. 22 are used in the construction of the output feature vector for the linear layer to perform the classification task. Specifically, the feature vector is generated by appending successively more and more central moments. We denote these appended feature vectors as $\vec{\mu}_{\leq p}$ for feature vectors containing up to p central moments, e.g.

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

is a feature vector constructed from appending the flattened covariance to the mean. The first-order moment here is a vector to denote 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

Supplementary Figure 11 Second-order central moment (covariance) of quantum reservoir output over spiral dataset. These correlation matrices were generated from calculating the covariance over measurement outcomes in a reservoir run, then averaged over the entire dataset.

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 $\binom{M+p-1}{p}$, 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 $\dim(\vec{\mu}_{\leq 3}) = 8 + 36 + 120 = 164$. This output dimension for the results presented in the main text is further reduced as discussed in the following paragraphs below.

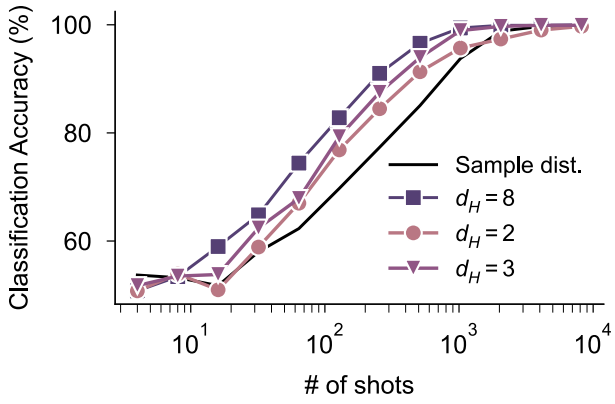


Supplementary Figure 12 Third-order central moment of quantum reservoir output over spiral dataset. These third-order central moments are plotted as an array of 2D matrices, such that the i-th column corresponds to the 2D matrix μ_{ijk}. The two rows corresponds to the two different classes of signals. In the third order, one can begin to see differences between the two classes by eye.

Supplementary Figure 10 contains classification results on the spiral dataset (Fig. 2) as a function of the number of shots for the feature vectors $\vec{\mu}_1, \vec{\mu}_{\leq 2}$ and $\vec{\mu}_{\leq 3}$. We see that our quantum reservoir has non-trivial third-order correlations and that the reservoir leverages these correlations to boost classification accuracy. The covariance matrix averaged over the entire spiral dataset is plotted in Supplementary Figure 11, and the third-order correlations are plotted in Supplementary Figure 12 – plotted as a set of 2D matrices. In the third-order correlations in particular, we can begin to pick out by eye the differences in the two classes.

This construction generally allows us to construct feature vectors that are smaller than the probability distribution over all possible measurement trajectories, which is 2^M dimensional. However, as can be seen in Supplementary Figure 11, there is yet redundant information even after taking only the symmetric part - specifically, that the information tends to be very local and that measurements far apart tend not to be correlated. This has the physical interpretation that while measurements are indeed correlated, even possessing higher-order correlations, this correlation tends to be local due to the finite memory of the system. This motivates us to further restrict our feature vector to only capture the essential local correlations.

Supplementary Figure 13 compares the classification performance of feature vectors generated with up to third-order moments, where we truncate the locality of the correlations. That is, the elements of the third order central moment $(\mu_3)_{ijk}$ are set to zero if $|i-j| > d_H$ or $|i-k| > d_H$, for some integer d_H. We note that including third-order correlations between measurements that are up to three 'sites' away nearly reproduces the classification accuracy of when you include all third-order central moments. Additionally, we compared the construction of feature vectors using truncated moments up to third-order with that of using the full sampled distribution as the feature vector and found that the former performed much better (Supplementary Figure 10). These last two statements were found to be true for all tasks presented in this paper. For $M = 8$ measurements, the truncation of long range correlations further reduces the output feature size from 164 down to 94.



Supplementary Figure 13 Spiral classification accuracy as a function of locality of up-to third-order correlations Classification accuracy for the spiral task using up to third-order central moments truncating correlations to only include correlations up to d_H. We find that we can achieve high accuracy using local third-order correlations, saturating the accuracy when only keeping correlators up to $d_H \leq 3$. The performance of using the sample distribution is also compared which performs worse than using the central moments as output features, despite containing more information (see text).

Training the linear layer

The only component of the reservoir that was trained to fit the dataset processed by the reservoir was the linear layer applied to the feature that the physical reservoir produced. The linear layer was an $R \times C$ matrix W_{train} and C-dimensional vector v_{train} applied to the R-dimensional reservoir feature x to get

$$y = x^T W_{\text{train}} + v_{\text{train}} \tag{23}$$

where the largest of the C elements of y corresponded to the predicted class of the data point (C is the number of classes). To train the linear layer, we chose between two different approaches: the pseudo-inverse method and back-propagation through a softmax function on the output. 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 there is not a perfect correspondence between our loss and the classification inaccuracy.

First, we will describe the pseudo-inverse method. Let X be a $N \times (R+1)$ matrix consisting of R-dimensional reservoir features generated for N training points, with a column of 1's appended (this is to compute both W_{train} and v_{train} at once). Let Y be an $N \times C$ matrix consisting of $C-$dimensional row vectors that serve as labels for the training points such that $Y_{i,j} = 1$ if j corresponds to the class of the i^{th} training point and zero otherwise. For an $\epsilon > 0$, we construct W'_{train} (W_{train} appended with v_{train}) as :

$$W'_{\text{train}} = \left(X^T X + \epsilon I\right)^{-1} X^T Y \tag{24}$$

In our case, the value of ϵ was swept to maximize the accuracy of the classification. In the limit of $\epsilon \to 0$, the pseudo-inverse matrix of Eq. 24 is provably optimal for minimizing $||XW_{\text{train}} - Y||_2^2$ [11] up to numerical stability, and so has been a popular choice for training the linear layer at the output of reservoirs [12–15]. However, our goal was to classify the input signals based on the *largest* element of the final output vector. Consequently, the linear layer that resulted in the lowest mean squared error with our labels was not always the linear layer that gave us the best accuracy.

For this reason, we also used a second training method for our linear layer. This approach used softmax, a popular choice for classifiers in neural networks [16] and back-propagation using the automatic differentiation package from PyTorch [17]. 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. In this approach, the prediction vector y from Eq. 23 is passed through the "Softmax" activation function:

$$(y_{\text{prediction}})_i = \frac{\exp(y_i)}{\sum_{j=1}^{C} \exp(y_j)} \tag{25}$$

We then computed the mean squared error between the resulting $y_{\text{prediction}}$ and the label for the training point that produced the underlying reservoir feature x. Finally, we used back-propagation to compute the gradient for our linear layer. The linear layer was then updated using the ADAM optimizer [18] with the default settings of $\beta_1 = 0.9, \beta_2 = 0.999$ and a learning rate of 0.01. For our reservoirs, we tried both methods of training the linear layer and used whichever yielded the best accuracies. Empirically, we found that while pseudo-inverse training was better in some cases, training the linear layer with back-propagation often yielded quite large accuracy advantages over pseudo-inverse.

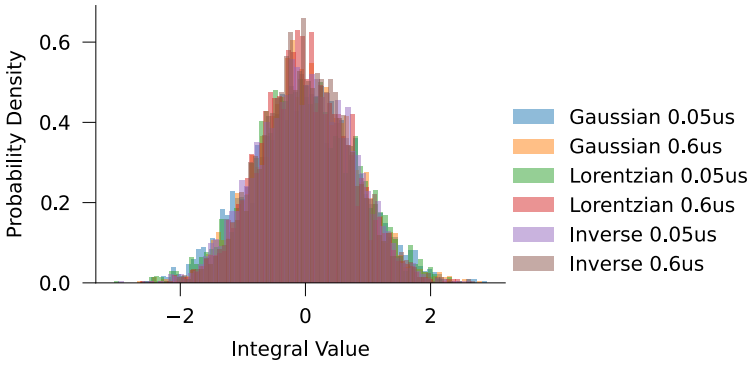
Supplementary Note 5 – Supplementary information machine learning tasks

Classification of Radio-Frequency signals

In this section, we discuss about the algorithm for generating the dataset for the classification of digital modulation schemes on radio signals. The digital modulation scheme involves encoding sequences of binary values into the amplitude and phase of a radio signal for a fixed duration. The number of binary values encoded depends on the modulation scheme. For example, for BPSK (binary phase shift key), each symbol (change in property of the signal) encodes one bit of information. For 32QAM (quadrature amplitude key), there are 32 possible values, which allows each symbol to contain 5 bits of information. For this task, we keep the symbol rate fixed across all the tasks. Moreover, the pulses generated by the arbitrary waveform generator (AWG) all occur at the baseband frequency. This signal is then upconverted to the frequency of the cavity before sent to the device. To generate the set of possible sequences, we randomly select each symbol with equal probability. This corresponds to the case of each possible binary string of digits encoded to be equally likely. Due to memory constraints on the AWG, we cannot output a continuous encoded signal for long durations, corresponding to the regime of large samples of the reservoir. We circumvent this constraint by realizing that, for this task, there are no correlations in the encoded binary digit sequence (since each symbol is equally likely). Therefore, the probability of a long binary digit sequence can be correctly emulated by sampling multiple short binary digit sequences and concatenating them together. For this task, we can simply achieve this by generating a signal with eight symbols, which is the number of symbols enter our QRC before its state is completely reset.

Classification of noisy signals

To generate the dataset describing the task of classifying noisy signals using the QRC (see Fig. 4), we start with emulating white noise. At each time step of the sampling rate of the AWG, we choose a value for in-phase and quadrature signals uniformly between the unit interval (up to an overall normalization). While this is limited to the sampling rate of the AWG (around 2×10^9 samples per second), this is much larger than any relevant time scale of the experiment. Therefore the approximation of broadband white noise is appropriate to describe the effect of the signal on the system. We then apply "kernels" as a convolution in the time domain to each need seed of the white noise generated signal. This can also be thought of a bandpass filtering function in frequency domain. The classification task is then to identify the kernel. Each kernel is defined by a time domain function. The only hyper-parameter to describe each kernel is the overall scaling value. In this work, we set the DC component of this kernel

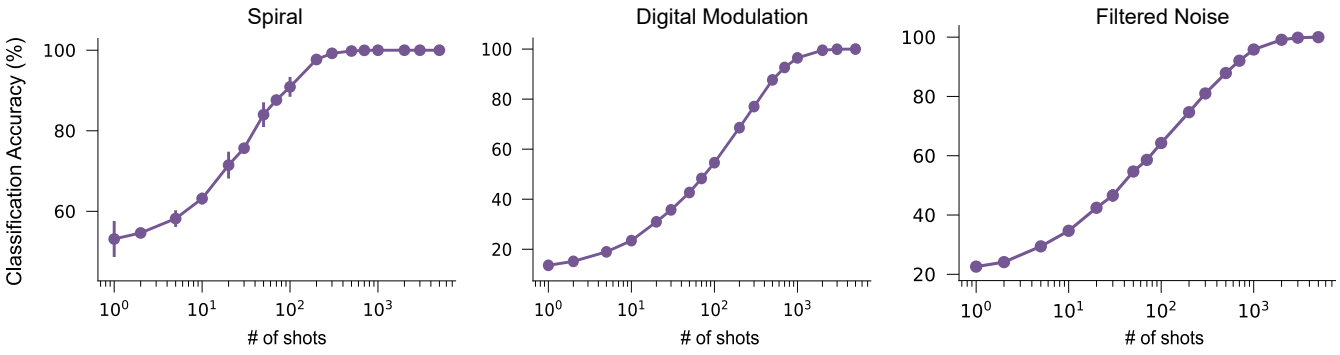


Supplementary Figure 14 Histogram of integrated value of the input signal for the noise classification task. In this work, we enforced a normalization condition on the amplitude of the filter functions. We set the normalization such that the long time integral of the signal corresponds to a value with zero mean and the same standard deviation. This is visually seen from the probability density function from the dataset of the classes of noisy signals. We do this to ensure that any reservoir which simply integrated the signal, before applying a non linear kernel, cannot classify the different signals. The fact that our reservoir is able to solve this task can therefore be associated with the continuous-time processing by the cross-Kerr interaction between the qubit and the cavity. Enforcing this normalization is mathematically equivalent to setting the DC component of the filter functions to be the same value in frequency space.

in frequency domain to be the same for all classes (set to unit value without loss of generality). In the time domain, this corresponds to scaling the amplitude such that the area enclosed by the filter function in time is the same for all functions.. We do this to make sure that a direct integration of the signal over a time domain much longer than the correlation length introduced by the kernel, cannot distinguish the signals from each other (see Supplementary Figure 14). The above normalization ensures the random variable associated with this integrated value is the same for all distributions. Therefore, this ensures that any ability of the reservoir to classify the signals arises intrinsically from its computational capacity to distinguish short-time correlations (in this work we choose a correlation time scale of 50ns and 600ns, with kernel functions of Gaussian, Lorentzian, and the Inverse function: generating a total of 6 classes.).

Supplementary Note 6 – Simulation of the quantum reservoir

Introduction



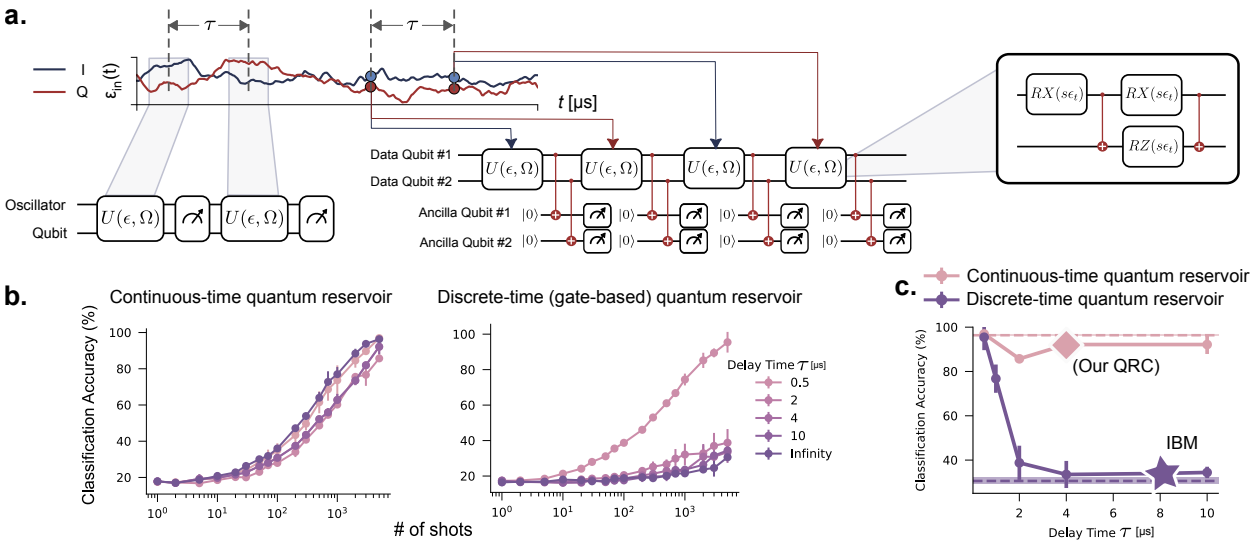
Supplementary Figure 15 Classification accuracies obtained from numerical simulations of the QRC. Numerical simulations of the quantum reservoir can help guide the expected performance of the system in experiment. Here, we simulate the three main tasks considered in the paper: Spiral (a time-independent input with the goal of classifying the arms of the spiral), Digital Modulation (a slow-varying, time-dependent signal with the goal of identifying the digital modulation scheme used in the signal), Correlated Noise (a fast-varying, time-dependent signal with the goal of classifying the "kernel" of the correlation function).

Classical simulations of the QRC can provide insight into the expected computational capacity in experiment. For our work, classical simulations of the dynamics of the reservoir were primarily performed with the aid of QuTiP [19]. The algorithm to estimate the classification accuracy for a given task then follows the same technique used in experiment, with training and testing datasets on the measurement outcomes of the simulation. We implement the Hamiltonian in Eq. 1 of the main text, by approximating the transmon as a qubit, and introducing a finite dimensional Fock truncation to the cavity subspace. It is important to ensure that the Fock truncation does not introduce any spurious effects, for it can be a source of non-physical non-linearities in the system. For example, a linear cavity, treated as a harmonic oscillator, only performs a linear transformation on an incoming analog radio frequency signal. However, if in simulation, the support of the state of the cavity exceeds the Fock truncation of the simulation, numerical errors introduce non-Gaussian states in the cavity mode. Such effects will depend non-linearly on the input, and hence can effectively act as a "good" (but of course unphysical) reservoir! To ensure this doesn't happen in simulation, at every step of the unitary evolution, we monitor the probability of the wavefunction on the largest Fock state in the simulation. If this value goes above 1% during the simulation, a warning is raised, and the results of the simulations are discarded.

To make the simulations efficient, we make certain assumptions on the quantum system. Firstly, we treat the reservoir controls of qubit rotations and conditional displacements with a "gate"-based unitary. However, to take into account the analog, continuous dynamical evolution implemented by the cross-Kerr interaction term in the Hamiltonian, the interval of the input into the system is implemented with the full time-dependent Hamiltonian evolution (using QuTiP's "mesolve" functions). Finally, another approximation we make (in favor of simulation speed) is ignoring decoherence effects. To ensure this approximation is valid, we performed simulation with the stochastic wavefunction approach with photon loss and qubit dephasing rates measured in experiments [20]. We obtain differences in expected classification accuracies within error bars (which are obtained from different datasets from repeated simulations). This also gives us confidence that the role of decoherence in the system plays a minimal role in the computational capacity of the reservoir. The results of the simulations with the third central moment are plotted in Supplementary Figure 15. Interestingly, the performance as a function of the number of samples agrees to experiment within the same order of magnitude. This gives a good estimate for the experimental time required to produce a classification accuracy versus shots curves in experiments. For all three tasks, the reservoir approaches 100% accuracy with sufficient samples or integration time of the input.

The advantage of continuous-time continuous-variable QRCs over discrete-time qubit-based QRCs

In this section, we benchmark the performance of our continuous-time-continuous variable QRC in comparison with other hardware implementations of reservoirs. To highlight the benefit of our QRC in processing time varying input signals, we compare the simulation of our reservoir with that of a recent QRC scheme involving repeated measurements on a multi-qubit based superconducting circuit quantum system [21]. For this comparison, we simulate the expected performance of both systems on the task of classifying different noise signals with the classes described in Fig. 4. While our reservoir can naturally interface with analog signal, this is not the case with the protocol introduced in [21]. For this simulation, the signal is sampled at discrete times and input to the system as a scalar parameter (one for the in-phase value and one for quadrature value). To highlight the advantage of our QRC, we slightly modify the task introduced in the Fig. 4 . Here, we normalize the six filter functions such that the integral of the filter function in frequency domain is kept constant. We do this such that the standard deviation associated with the distribution of the sampled signal is the same across all signals. The only information distinguishing the signals is in the correlation between two close samples in time. To elucidate this reasoning, we simulate the performance of the two reservoirs as a function of the time duration in between two samples of the signal (in the case of the discrete qubit based reservoir) and integration windows (for our analog reservoir). Such a finite duration can arise from finite-pulse durations of reservoir protocols, qubit-measurement times, and the finite latency introduced by the classical FPGA processor. For example, for our experiment, this time is around $4\mu s$, mostly arising from the measurement of the qubit and the parity of the cavity. In experiment (Fig. 4 of the main text), we had generated and timed the input wave forms such that the delay between inputs is essentially $0\mu s$. For a typical IBM quantum device with mid-circuit measurement, the protocol used in Ref. [21], the finite latency can be estimated to be around $8\mu s$ [22]. The protocol for the discrete-time quantum reservoir is designed to only act on real-valued input signals. However, for a continuous signal in the rotating frame, we have both the in-phase and quadrature values. For the experimental quadrature, these values correspond to displacements on the oscillator in orthogonal directions. To

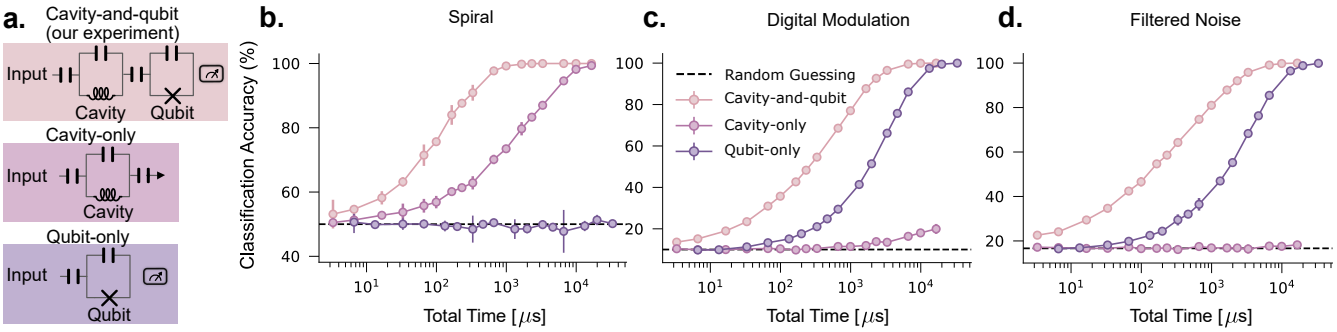


Supplementary Figure 16 Simulations comparing the performance of a continuous-time quantum reservoir (our work), and a discrete-time qubit-based quantum reservoir, based on a recently implemented protocol [21]. (a.) Schematic reservoir protocol for our experiment (on the left), versus that introduced in [21], involving two qubits as the data qubits, and two as the ancillas. The reservoir consists of single qubit rotations interleaved with CNOT gates between the data qubits and data and ancilla qubits. In this simulation, we simulate the performance of the two reservoirs as a function of the delay τ between two durations of the input. Such a delay can be introduced by the finite pulse durations and latencies introduced by the FPGA. **(b.)** Classification accuracy curves for the two reservoirs as a function of shots, for different values of delays. The continuous time analog reservoir is much more robust to delays between inputs compared to the discrete time reservoir implementation. The errorbars here indicate the variation in testing accuracy over the testing dataset. **(c.)** Plot of accuracies at 5000 shots for the two reservoir implementations as a function of delay time. Experimentally relevant times include $4\mu s$ for our experiment, and around $8\mu s$ [22] for the experimental realization of [21] on an IBM quantum computer.

extend the scheme presented in [21], we do the following minimal change: we interleave the between sample points of in-phase and quadrature values. We could have chosen these points with the delay of τ in between each. However, this might have had the effect of introducing twice the delay compared to the continuous time reservoir. Therefore, we chose the relaxed constraint of the input such that both the in-phase and quadrature values are chosen at the same point, with just a delay in between two different in-phase and quadrature points.

Comparison to other reservoirs

A cavity coupled to a qubit is a hardware efficient quantum system to perform reservoir computing on analog signals. In this section, we motivate this by simulating the performance of other natural choices of quantum reservoirs: a single qubit, and a single cavity. The protocols for these systems are inspired by what one can naturally perform in experiment. To make a reservoir with a cavity, we couple the input into the cavity (as is the case for the experimental design). To readout the cavity, we perform a transmission style Homodyne measurement, which infers the mean field value of the cavity. This is a continuous form of measurement, where the output feature is a time dependent radio frequency signal at the frequency of the cavity mode. Since the cavity is always in a coherent state, the output time trace is linearly dependent on the incoming signal. For a fair comparison, we only use a handful of values from the time trace (as many as the number of measurements in the experiment). While this might seem restrictive, we process this via the same method as the case of the experimental reservoir, by computing the functional definition of the central moments. This does not necessarily make sense for this protocol, since the outputs do not correspond to samples from a discrete probability distribution, but can nevertheless introduce non-linearities in the representation of the feature vector. These non-linearities can improve the performance of the reservoir beyond a linear layer. This is observed for the case of time-independent Spiral classification, where the cavity reservoir performs better than random. This performance is solely due to the "post-processing" of the output of the reservoir we adopt for our experiment. However, for time dependent tasks, the performance is hardly better than random.

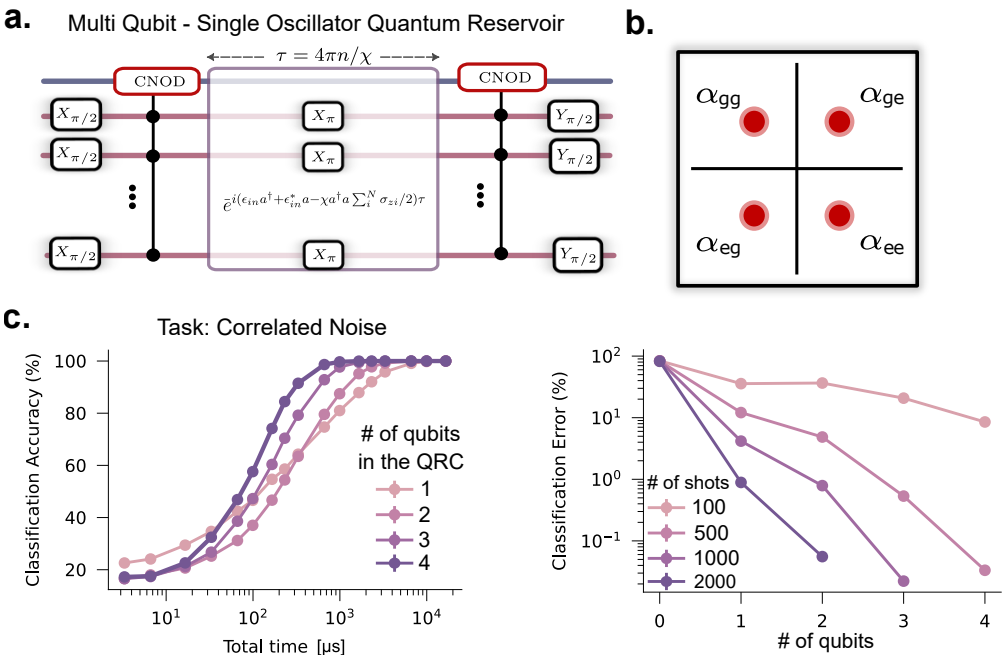


Supplementary Figure 17 Simulations of the performance of the cavity-and-qubit reservoir system with its components: just a cavity reservoir and just a qubit reservoir. (a.) The circuit diagrams of the three quantum reservoir simulated here. 1) The circuit diagram of the experimental QRC in this work: a cavity coupled to a qubit. The input is interfaced to the cavity, and the state of the qubit is measured using the standard dispersive readout technique with a readout resonator. 2) The circuit diagram of a cavity as a reservoir. In this case, the natural output of the cavity is the transmitted signal. 3) The circuit diagram of a qubit reservoir. In this case the input is coupled instead to the qubit. **(b.)** Simulation of the classification accuracy of the three different reservoirs for the three tasks considered in the paper. As a fair comparison, the output feature vector dimension is kept constant for all three reservoir. The experimental setup drastically outperforms either of its components, just a cavity (in **(c.)** and just a qubit (in **(d.)**), which highlights the important role of entanglement in classification accuracy. Errorbars in the classification accuracy reflect the variation in accuracy over the testing dataset.

Another natural candidate is a single qubit reservoir. For this case, we directly interface the signal to the qubit. A qubit is able to naturally represent non-linear functions of the input, which can be intuitively seen by visualizing the action of a qubit rotation on a Bloch sphere. As a fair comparison, we choose the same qubit reservoir controls in experiment, which involve qubit pulses before, during and after the continuous input. The output is a string of binary outcomes of qubit measurements, which can be done experimentally with the use of a readout resonator. Each reservoir of the qubit lasts twice as long as the experimental QRC to obtain the same feature vector size. This is then processed the same way as the cavity-qubit coupled reservoir, before applying a trained linear layer. Interestingly, the qubit fails to perform better than random for the Spiral task. On the other hand, it is able to reach near 100% accuracy for the time dependent signal classification tasks. The ability for even a single qubit to successfully perform a many-class classification task is illuminating at the remarkable processing capabilities of reservoir. However the total input signal required can be more than order of magnitude longer compared to that for the experimental QRC to achieve the same accuracy. The ability of a cavity coupled to a qubit system to perform significantly better than either of its components provides a clear picture of the important role entanglement can play.

Multi-qubit reservoirs

Quantum reservoir computing is a promising paradigm in the NISQ era. It is therefore interesting to consider the potential benefits in performance with larger devices, which are within reach of today's experimental capabilities. As a natural extension of our quantum reservoir, we consider a scenario of one continuous variable cavity mode dispersively coupled to multiple qubits. For simplicity, we assume the dispersive strength of each qubit to the cavity is the same. To motivate the capacity of such a reservoir, we simulate the system for up to four qubits to estimate the classification accuracy for the task of identifying correlated noise signals. The unitary protocol is illustrated in Fig 18 (a.). The protocol begins with a $\pi/2$ pulse on each qubit, which brings the state of the qubit onto the equator of the Bloch sphere. To entangle the qubits with cavity, we simulate the action of a generalized, multi-qubit-conditioned cavity displacement. This involves a displacement on the cavity, whose value is different for each of 2^N N-qubit possibilities. An example of the action of this operator is depicted in Fig 18 (b.) for the case of two qubits. Here, there are four possible qubit states, and each is associated with a displacement value of the four corners of the square. The displacement values were chosen somewhat arbitrarily, but serve to illustrate an efficient multi-component entanglement. For the case of two qubits, the real and imaginary components of the displacement where either ± 0.5. For the case of three qubits, there are eight total possible states. The set of displacements chosen form a three-by-three grid state, ranging between ± 1, excluding the center of this grid (which is centered at the origin). For the case of four qubits, a four-by-four grid uniformly distributed between ± 1.5 covers all sixteen possibilities.



Supplementary Figure 18 Exploring the computational capacity of more complex reservoirs, involving a single cavity mode coupled to multiple qubits. (a.) The reservoir protocol, before measurement. The protocol is inspired by naturally extending the protocol of the experiment. The state of each qubit, along with the parity of the cavity, is sampled afterwards. **(b.)** Schematic illustration of the state of the cavity after the generalized qubits conditioned cavity displacement implemented in this reservoir, for the case of two qubits. Here the cavity is displaced by a unique value of each 2^N combinations of qubit possibilities, for a reservoir with N qubits. **(c.)** To illustrate the computational capacity of such reservoir, we simulate the quantum system to obtain the classification accuracy as a function of the duration of signal received for the task of classifying correlated noise signals. Increasing the number of qubits generally increases the performance of the reservoir. For different values of shots of the reservoir, which corresponds a total time of input, we plot the classification error as a function of qubits. The case of zero qubits in the reservoir corresponds to just a cavity reservoir.

The correspondence between qubit states and displacements was somewhat arbitrary - the motivation is that even without much design choice, a reservoir can successfully implement machine learning! For these simulations, each position of the grid is associated with a decimal value, increasing sequentially from left to right, starting from the top left and progressing towards the bottom right (starting with zero). This decimal value is the decimal representation of the qubit-state bit string that the displacement is conditioned on.

After the multi-qubit entangling conditional displacement gate, the cavity is subject to the input. The dynamics of the system are influenced by the cavity-qubit coupled dispersive interactions, where the interaction strength is the same between the cavity and all qubits and set to that of the experiment. Like the experimental QRC, each qubit is flipped with a π pulse in the middle of the input. The protocol ends with the same conditional displacement, before a $\pi/2$ pulse. The output of the reservoir is the measurement of each qubit, along with the parity measurement of the cavity. This protocol is repeated four times, to match the experimental protocol as much as possible.

Fig 18 (c.) is the classification accuracy as a function of the total time of input signal received for the reservoirs with different number of qubits, for the task of noise classification. While they all achieve essentially 100% accuracy, the total time required to achieve this accuracy drops significantly with increasing number of qubits. Other than the single qubit reservoir (the experimental protocol), the performance of the reservoir is similar both at the low and high signal duration regime, differing only in the intermediate regime. The reason for the difference in behavior of the performance for the case of cavity coupled to a single qubit is the slight change in reservoir protocol. To accurately account for the experimental protocol, the state of the qubit is determined by the outcome of the parity measurement of the cavity. This was not implemented in the simulations for multiple qubits. This ends up improving the performance of this reservoir for this task in the low signal duration regime. However, in the higher signal duration regime, increasing the number of qubits increases the accuracy.

The classification error as a function of number of qubits in the reservoir is plotted in Fig 18 (d.), including the case for just a cavity (zero number of qubits in the reservoir), for a select number of total shots of the entire reservoir. Very crudely, the error in classification seems to reduce exponentially with every additional qubit in the reservoir.

Supplementary Note 7 – Theoretical analysis of the expressivity of our QRC for time-independent signals

The ability of the QRC to perform better than an optimal linear layer on the input lies in the reservoir's ability to express many non-linear functions of the input—its expressivity. Here, we quantitatively characterize the class of functions which can be represented by the oscillator component of the QRC for a time-independent input. In this regime, the input can be represented by two variables: the values of the in-phase and quadrature components. The output feature vector from the QRC is then a function of these two variables.

We denote $\alpha = |\alpha|e^{i\phi_\alpha}$, $\beta = |\beta|e^{i\phi_\beta}$, and set $|\alpha| = 1/2$. Choosing different values of ϕ_α gives rise to different output features of the QRC. In this experiment, we pick $\phi_\alpha \in \{0, \pi/2\}$, but in principle, one can add to the feature vector with more choices of ϕ_α. For example, one can choose $\phi_\alpha \in \{0, \omega, 2\omega, \ldots, (r-1)\omega\}$ where $\omega = \frac{2\pi}{r}$. The final output after the linear layer is an arbitrary linear combination of all the $p_\alpha(\beta)$ functions.

Intuitively, the larger r, the more expressive the function space spanned by these features. Furthermore, the higher-order central moments allow the output feature vector to represent powers of this probability: $p_\alpha(\beta)^n$, for moments up-to the nth-order. We have shown that the qubit measurements extract the phase information of the input complex number β. Below we will focus on the oscillator parity measurement which is sensitive to the magnitude of β. Recall that the post-measurement (unnormalized) state of the cavity can be described by a sequence of alternating displacements and parity measurements (Eq. 14):

$$|\Psi_{\vec{x}}(\beta)\rangle = P_{x_M} D(\beta) \cdots P_{x_2} D(\beta) P_{x_1} D(\beta)|0\rangle, \tag{26}$$

where P_{x_i} is the projector of the i-th parity measurement with outcome $x_i \in \{0, 1\}$, with '0' standing for 'even' and '1' for 'odd'. That is, $P_{x_i} = \frac{I+(-1)^{x_i}\Pi}{2}$, where $\Pi = (-1)^{a^\dagger a}$. The corresponding probability of obtaining $\vec{x} = (x_1, x_2, \ldots, x_M)$ as the sequence of measurement results given the input β is

$$\Pr[\vec{x}|\beta] = \langle\Psi_{\vec{x}}(\beta)|\Psi_{\vec{x}}(\beta)\rangle. \tag{27}$$

To obtain a simplified expression for $\Pr[\vec{x}|\beta]$, we will make use of the following formula:

$$P_x D(\beta) P_y = \frac{D(\beta) + (-1)^{x\oplus y} D(-\beta)}{2} P_y, \quad \forall x \in \{0,1\}, \forall y \in \{0,1\}, \tag{28}$$

which is an easy application of the commutation relation $\Pi D(\beta) = D(-\beta)\Pi$, with the latter being derived from $\Pi a = -a\Pi$. Using Eq. 28, we can remove all the explicit parity projectors in Eq. 26:

$$|\Psi_{\vec{x}}(\beta)\rangle = \left(\prod_{i=1}^{M} \frac{D(\beta) + (-1)^{x_i \oplus x_{i-1}} D(-\beta)}{2}\right)|0\rangle\,, \tag{29}$$

where for notational simplicity we have prepended the bit-string $\vec{x}$ by $x_0 \equiv 0$. Note that the order of the product does not matter since the terms commute with each other. It follows that:

$$\begin{aligned}\Pr[\vec{x}|\beta] &= \langle 0|\left(\prod_{i=1}^{M} \frac{D(-\beta) + (-1)^{x_i \oplus x_{i-1}} D(\beta)}{2}\right)\left(\prod_{i=1}^{M} \frac{D(\beta) + (-1)^{x_i \oplus x_{i-1}} D(-\beta)}{2}\right)|0\rangle \\ &= \langle 0|\left(\prod_{i=1}^{M} \left[\frac{1}{2} + (-1)^{x_i \oplus x_{i-1}} \frac{D(2\beta) + D(-2\beta)}{4}\right]\right)|0\rangle.\end{aligned} \tag{30}$$

There are multiple methods to encode the measurements of the QRC. Representing every binary string of measurement outcomes as the feature, the output of the QRC are all the probabilities $\{\Pr[\vec{x}|\beta]\}_{\vec{x}\in\{0,1\}^M}$. From

Eq. 30, it is not hard to see that when regarded as functions of β, these 2^M features linearly span a $(M+1)$-dimensional function space that has the following basis functions:

$$f_k(\beta) := \langle 0| D(2k\beta) |0\rangle = e^{-2k^2|\beta|^2}, \quad k = 0, 1, 2, \ldots, M. \tag{31}$$

Therefore, the set of all functions realizable by the QRC combined with the linear layer is

$$\{c_0 f_0(\beta) + c_1 f_1(\beta) + \cdots + c_M f_M(\beta) : c_0, c_1, \ldots, c_M \in \mathbb{R}\}. \tag{32}$$

Given the large redundancy of the output feature encoding manifested above, a compact representation can be the centralized moments $\mu_{i_1,i_2,\ldots,i_k}(\beta) := \mathbb{E}\left[(x_{i_1} - \mathbb{E}[x_{i_1}])\,(x_{i_2} - \mathbb{E}[x_{i_2}]) \cdots (x_{i_k} - \mathbb{E}[x_{i_k}])\right]$. These feature functions contain terms like $\mathbb{E}[x_1]\,\mathbb{E}[x_2]$, $\mathbb{E}[x_1]^2$, $\mathbb{E}[x_1]\,\mathbb{E}[x_2]\,\mathbb{E}[x_3]$, and so on. In particular, for any k, $\mathbb{E}[x_1]^k = \left(\frac{1}{2} - \frac{e^{-2|\beta|^2}}{2}\right)^k$ can be written as a linear combination of centralized moments of order less than or equal to k. It follows that the QRC using at most k-th order centralized moments combined with the linear layer can realize (but not limited to) the following vector space of functions:

$$\mathcal{H}_{\text{parity}} := \left\{c_0 + c_1 e^{-2|\beta|^2} + c_2 \left(e^{-2|\beta|^2}\right)^2 + \cdots + c_k \left(e^{-2|\beta|^2}\right)^k : c_0, c_1, \ldots, c_k \in \mathbb{R}\right\}. \tag{33}$$

Note that $\mathcal{H}_{\text{parity}}$ is exactly the set of all degree-k polynomials in the variable $w \equiv e^{-2|\beta|^2}$. Suppose that in some classification task, the magnitude of the input has an upper bound, say, $|\beta| \leq 1$, then w takes value in the closed interval $[e^{-2}, 1]$. By the Stone–Weierstrass theorem, in the limit $k \to \infty$, $\mathcal{H}_k$ approximates all continuous functions of w on $[e^{-2}, 1]$, and hence all continuous functions of $|\beta|$ on $[0, 1]$.

Supplementary Note 8 – Leaky Echo State Networks (LESN)

Background

Leaky echo state networks [23] are a generalization of echo state networks (ESN) [24] that were found to outperform their parent design in prediction and classification of slow dynamic systems, noisy time series and time-warped dynamic patterns [13]. Given a sequence of inputs $\{u_n\}_{n=1}^N, u_n \in \mathbb{R}^D$, the state of the LESN reservoir after the n^{th} input u_n, x_n, is given by the following equation:

$$x_n = (1 - a\gamma)x_{n-1} + \gamma f\left(W_{\text{in}} u_n + W_{\text{res}} x_{n-1}\right). \tag{34}$$

Here, a, γ are fixed hyper-parameters in $[0, 1]$, and f is a nonlinear activation function. W_{in} is the $R \times D$ "encoding" matrix whose elements are selected uniformly at random from the interval $[-w_{in}, w_{in}]$, where D is the dimension of the input, R is the dimension of the reservoir, and w_{in} is a fixed hyper-parameter. W_{res} is the $R \times R$ "reservoir" matrix. This matrix is constructed by first generating a matrix W_R, which is a random matrix whose elements are chosen to be zero with probability $1 - p_s$ and a number sampled uniformly from the interval $[-1, 1]$ with probability p_s. The largest-magnitude singular value of this matrix, $\lambda_{\max}(W_R)$ is computed, and the reservoir matrix W_{res} is defined as:

$$W_{\text{res}} = \frac{\rho}{|\lambda_{\max}(W_R)|} W_R \tag{35}$$

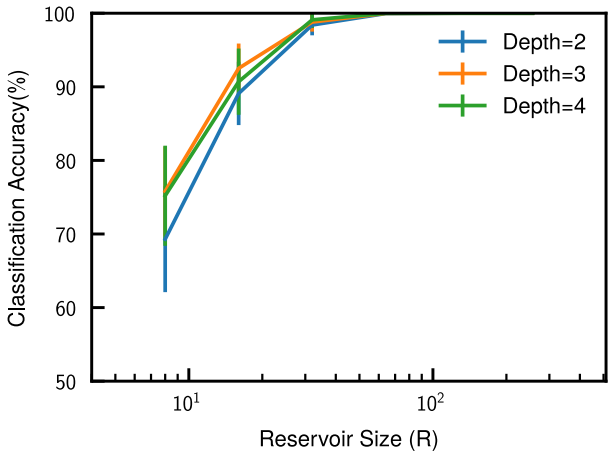
where ρ is a fixed scaling hyper-parameter. Finally, the n^{th} output of the reservoir, y_n, is given by

$$y_n = W_{\text{train}} x_n, \tag{36}$$

where W_{train} is a $C \times R$ trainable linear layer, where C is the dimension of the desired output vector.

Digital reservoir comparison

As a way of benchmarking the computational capacity of our physical reservoir, we compared it to the performance of a *digital* reservoir - an LESN - at varying widths and depths. We focused on the accuracy of classifying the spiral, since this is the most direct point of comparison, as the goal was to classify individual points of a signal



Supplementary Figure 19 Mean spiral classification accuracies and their standard deviations over 100 randomly generated LESN's.

rather, than multiple separate signals per-shot as with the time-dependent case. Here, for a depth of N, we sent in N identical two-dimensional data points (x, y) (so $D = 2$) corresponding to the I and Q components of the signal that our experimental reservoir is meant to process, i.e. the spiral point coordinates. We used the rectified linear unit (ReLU) as our nonlinear activation function. Traditionally, sigmoid or tanh activation functions are used for LESN's [13, 15, 23], but ReLU was found to work better for our application.

To investigate how "trivial" it was to generate a classifier with the same capacity as our experiment, we generated 100 such LESN's at random, and found their average performance, and standard deviation. Hyper-parameters $a, \gamma, w_{\text{in}}, \rho$, and sparsity (p_s) were tuned in sweeps to improve performance as much as possible for each width and depth, in order to give the digital reservoir a competitive chance. The reservoir's computational capacity varies by the number of shots. Comparing Fig. 2b to Fig 19, we found that, at around 10^3 shots, our physical reservoir achieved a performance comparable to that of about that of a 32-dimensional LESN reservoir, as seen by the fact that both oscillate around 99% classification accuracy, within about one percent. In Fig 19, a 64-dimensional reservoir was found to be enough to classify the spiral data points with perfect accuracy and a fairly wide choice of parameters. Our reservoir, then, achieved at *least* the capacity of a 64-dimensional LESN reservoir, past around $5 * 10^3$ shots.

In the conventional view of reservoirs, the data must be sent into into a higher dimensional space where linear separability becomes possible [13]. 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.

References

[1] Blais, A., Grimsmo, A.L., Girvin, S.M., Wallraff, A., Circuit quantum electrodynamics. *Reviews of Modern Physics* **93**(2) (2021)

[2] Chou, K.S., Teleported operations between logical qubits in circuit quantum electrodynamics. PhD thesis, Yale University (2018)

[3] Diringer, A.A., Blumenthal, E., Grinberg, A., Jiang, L., Hacohen-Gourgy, S., Conditional not displacement: fast multi-oscillator control with a single qubit (2023) arXiv:2301.09831

[4] Heeres, R.W., Vlastakis, B., Holland, E., Krastanov, S., Albert, V.V., Frunzio, L., Jiang, L., Schoelkopf, R.J., Cavity state manipulation using photon-number selective phase gates. *Physical Review Letters* **115**(13) (2015)

[5] Wang, S., Fontana, E., Cerezo, M., Sharma, K., Sone, A., Cincio, L., Coles, P.J., Noise-induced barren plateaus in variational quantum algorithms. *Nature communications* **12**(1), 6961 (2021)

[6] Royer, A., Wigner function as the expectation value of a parity operator. *Phys. Rev. A* **15**(2), 449–450 (1977)

[7] Gambetta, J., Blais, A., Schuster, D.I., Wallraff, A., Frunzio, L., Majer, J., Devoret, M.H., Girvin, S.M., Schoelkopf, R.J., Qubit-photon interactions in a cavity: Measurement-induced dephasing and number splitting. *Phys. Rev. A* **74**(4), 042318 (2006)

[8] Schuster, D.I., Houck, A.A., Schreier, J.A., Wallraff, A., Gambetta, J.M., Blais, A., Frunzio, L., Majer, J., Johnson, B., Devoret, M.H., et al., Resolving photon number states in a superconducting circuit. *Nature* **445**(7127), 515–518 (2007)

[9] Wang, C.S., Curtis, J.C., Lester, B.J., Zhang, Y., Gao, Y.Y., Freeze, J., Batista, V.S., Vaccaro, P.H., Chuang, I.L., Frunzio, L., et al., Efficient multiphoton sampling of molecular vibronic spectra on a superconducting bosonic processor. *Phys. Rev. X* **10**(2), 021060 (2020)

[10] Deng, X., Li, S., Chen, Z.-J., Ni, Z., Cai, Y., Mai, J., Zhang, L., Zheng, P., Yu, H., Zou, C.-L., et al., Heisenberg-limited quantum metrology using 100-photon fock states (2023) arXiv:2306.16919

[11] Björck, Å., Numerical Methods for Least Squares Problems. Society for Industrial and Applied Mathematics (1996). https://doi.org/10.1137/1.9781611971484

[12] Nakajima, K., Reservoir computing: theory, physical implementations, and applications. *IEICE Technical Report; IEICE Tech. Rep.* **118**(220), 149–154 (2018)

[13] Sun, C., Song, M., Cai, D., Zhang, B., Hong, S., Li, H., A systematic review of echo state networks from design to application. *IEEE Transactions on Artificial Intelligence*, 1–15 (2022)

[14] Gauthier, D.J., Bollt, E., Griffith, A., Barbosa, W.A.S., Next generation reservoir computing. *Nature Communications* **12**(1) (2021)

[15] Scardapane, S., Uncini, A., Semi-supervised echo state networks for audio classification. *Cognitive Computation* **9**, 125–135 (2017)

[16] Banerjee, K., Prasad C, V., Gupta, R.R., Vyas, K., H, A., Mishra, B., Exploring Alternatives to Softmax Function. *arXiv e-prints*, 2011–11538 (2020) arXiv:2011.11538 [cs.LG]

[17] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al., PyTorch: An Imperative Style, High-Performance Deep Learning Library. *arXiv e-prints*, 1912–01703 (2019)

[18] Kingma, D.P., Ba, J., Adam: A Method for Stochastic Optimization. *arXiv e-prints*, 1412–6980 (2014)

[19] Johansson, J.R., Nation, P.D., Nori, F., Qutip: An open-source python framework for the dynamics of open quantum systems. *Computer Physics Communications* **183**(8), 1760–1772 (2012)

[20] Dalibard, J., Castin, Y., Mølmer, K., Wave-function approach to dissipative processes in quantum optics. *Phys. Rev. Lett.* **68**(5), 580–583 (1992)

[21] Yasuda, T., Suzuki, Y., Kubota, T., Nakajima, K., Gao, Q., Zhang, W., Shimono, S., Nurdin, H.I., Yamamoto, N., Quantum reservoir computing with repeated measurements on superconducting devices (2023) arXiv:2310.06706

[22] Hua, F., Jin, Y., Chen, Y., Vittal, S., Krsulich, K., Bishop, L.S., Lapeyre, J., Javadi-Abhari, A., Zhang, E.Z., Exploiting qubit reuse through mid-circuit measurement and reset (2023) arXiv:2211.01925

[23] Jaeger, H., Lukoševičius, M., Popovici, D., Siewert, U., Optimization and applications of echo state networks with leaky-integrator neurons. *Neural Networks* **20**(3), 335–352 (2007). Echo State Networks and Liquid State Machines

[24] Jaeger, H., The "echo state" approach to analysing and training recurrent neural networks-with an erratum note. *Bonn, Germany: German National Research Center for Information Technology GMD Technical Report* **148**(34), 13 (2001)