Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Resource map of quantum reservoir computing against tuned classical baselines

Hikaru Wakaura

hikaruwakaura@gmail.com

QuantScape Inc

Research Article

Keywords: Quantum reservoir computing, quantum supremacy

Posted Date: November 2nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11548214/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Resource map of quantum reservoir computing against tuned classical baselines

Hikaru Wakaura[1, *]

[1] *QIRI (Quantum Integrated Research Institute Inc.), Tokyo 107-0061, Japan*

(Dated: October 28, 2026)

Quantum reservoir computing promises an exponentially large feature map at no training cost, but reported advantages rest on classical baselines of unmatched size or tuning, on exact expectation values, or on resource counts that omit measurement. Here we introduce a benchmark protocol that charges shots, experiment restarts and readout size together and apply it to 3451 exactly simulated cells (57 quantum reservoirs, 49 tasks) against three tuned classical reservoir families. No classical- or quantum-input task yields a win in both accuracy and resource scaling; the only winning cells are Forrelation estimators on oracle pairs drawn fresh at every step with no exposed structure, which beat every implemented classical estimator in every seed from field sizes of 16 thousand, with a matching classical query budget growing as the square root of the field size. Few-photon Kerr dynamics beats its mean-field twin only until the twin is tuned, spin readouts pay up to three orders of magnitude in shot noise where quadrature readouts pay threefold, and every structured oracle family falls to a classical shortcut. Quantum reservoir advantage therefore lives in the input access model, not in the reservoir; we give the superconducting-circuit parameters at which the protocol can be run.

I. INTRODUCTION

Physical reservoir computing uses the transient response of a dynamical system as a fixed nonlinear feature map and trains only a linear readout on it [1, 2]. The approach has been realised in optics, spintronics, mechanics and electronics, and it is attractive wherever training a deep network is too slow or too energy-hungry [3]. Quantum reservoir computing (QRC) carries the idea to quantum systems, where the number of observables grows exponentially with the number of qubits, so that a handful of spins could in principle furnish the feature space of a large classical network [4–6]. Experiments on nuclear spins, superconducting circuits, trapped neutral atoms and photonic chips now report competitive or superior accuracy on time-series benchmarks [7–11], and optimised small reservoirs are reported to beat tuned classical ones at equal readout cost [12].

Whether these results constitute a quantum advantage depends on the baseline. When a classical reservoir of the same size receives the same tuning effort, the gap shrinks or reverses [13–15]. When shot noise is included, a linear readout of raw inputs can outperform the quantum feature map at small reservoir sizes [16]. Theory identifies specific resources that classical linear reservoirs cannot reproduce, such as cross-time nonlinear products from Kerr feedback [17], non-Gaussian measurements [18], magic [19], and scrambling at the onset of chaos [20–22], and it also identifies the measurement cost that counterbalances them [23, 24]. Each claim concerns one reservoir family, one task family and one accounting convention, and the conventions disagree: some count readout weights, some physical nodes, some wall-clock time, and few count shots. However, no study has placed the candidate resources, the tasks and the costs on one map with a single accounting rule and symmetric tuning, so it remains unknown which computations, if any, let a quantum reservoir beat the best tuned classical model in both accuracy and resource scaling, and what those computations have in common.

Here we construct such a map. We simulate 57 quantum reservoir configurations exactly, spanning qubit-reset and Hamiltonian-encoded spin chains with z, xyz and two-body readouts, time multiplexing, intrinsic and engineered dissipation, online weak measurement, Gaussian and Kerr continuous-variable (CV) modes, quantum Kerr delay lines with feedback depth up to 100, a quantum-input reservoir, and a two-query Forrelation circuit used as an untrained feature map. We run them on 49 tasks (19 classical time series, 4 quantum-input and 26 oracle-input tasks), at shot counts from 10^2 to exact expectations, against echo-state networks, polynomial nonlinear autoregressive models and semiclassical Kerr delay lines, each tuned by a 20-draw random search at 22 sizes. We found that no classical-input cell wins against the best classical model once the classical side is tuned: over all 989 configuration–task pairs the quantum error is a median of 2.6 decades above the best classical error, and 0.8 decades for the best configuration per task; equal-readout-cost wins are small and are matched by an echo-state network of at most 100 nodes; few-photon Kerr dynamics beats its own mean field at identical parameters, but once the mean field is tuned the gain reverses on smooth tasks and shrinks to a tie on nonlinear-memory tasks; and quantum-input targets are easier for polynomial readouts of Pauli expectations. The result shows that the only cells winning on both axes are the Forrelation tasks whose oracle pair is drawn fresh at every step with no structure exposed to the querier, uniformly random pairs and Maiorana–McFarland pairs masked by a fresh random affine map: there the quantum estimator at 200 queries beats every implemented classical estimator

* h.wakaura@qiri.co.jp

with ten times the queries in every seed from $N = 2^{16}$ and 2^{14} respectively, and the matching classical budget grows with N. Every oracle family with exposed structure admits a classical shortcut that removes the win: a fixed public Kloosterman pair falls to an estimator that uses the definition of f, quadratic Gold pairs to a degree-2 reconstruction, unmasked Maiorana–McFarland pairs to an adaptive attack with at most $k^2 + 4k + 2$ queries on $n = 2k$ bits, and the drift tasks to a monitor that knows the ideal relation. The common denominator of quantum reservoir advantage is therefore the access model of the input, not a property of the reservoir.

II. METHODS

A. Accounting rule

Every cell of the map is one (quantum configuration, task, seed) triple. The quantum reservoir produces F_q feature time series $X \in \mathbb{R}^{T\times F_q}$ of exact expectation values together with their single-shot variances V. A ridge readout with validation-tuned regularisation (12 values from 10^{-10} to 10, selected on the last 20 % of the training segment) maps features to targets; the test error is the normalised mean-square error $\mathrm{NMSE} = \sum_t (y_t - \hat{y}_t)^2 / \sum_t (y_t - \bar{y})^2$, averaged over target components. Finite-shot readout with S shots per feature and time step is emulated by adding independent Gaussian noise of variance V/S to every feature, train and test alike, for $S \in \{10^2, \ldots, 10^6\}$ and $S = \infty$; $q(S)$ denotes the resulting error and $q \equiv q(\infty)$. This model charges S shots per feature but not the number of measurement settings (three bases for xyz readouts, V separate runs for time multiplexing) and ignores the correlation between features read from the same shots, so it undercharges the quantum side; the replay length $W = 30$ is a fixed convention, not a verified fading-memory cutoff (Appendix E). A projective readout destroys the reservoir state, so each time step must be replayed from a washout of $W = 30$ steps; an online weak-measurement readout or a memoryless circuit needs no replay, $W = 1$. The experimental cost of a cell is therefore $S\,W$ *shot-operations* per time step, and the budget used for the scaling criterion is 3×10^5 shot-operations, i.e. $S = 10^4$ for projective and $S = 10^5$ for online readouts.

Three classical families are tuned per (task, seed) and per size by a 20-draw random search with validation selection (a 50-draw budget strengthens the ESN by up to 14-fold on smooth tasks and changes no conclusion; Appendix C). The quantum configurations are fixed settings except the two tuned Kerr delay lines, which receive the same 20-draw search; the tuning asymmetry therefore favours the classical side: leaky echo-state networks (ESN) with $M \in \{3, 5, 10, 30, 100, 300, 1000\}$ nodes (spectral radius, input scale and leak tuned), polynomial nonlinear autoregressive models (poly-NARX; ridge on monomials of degree $p \le 3$ in $L \le 30$ lagged inputs, 11 sizes), and semiclassical Kerr delay-line reservoirs with $D \in \{10, 30, 100, 300\}$ virtual nodes (coherent amplitudes $a \to \eta e^{i\chi|a|^2}(\mathrm{shift}(a) + \varepsilon u\, m)$, three features per node; Kerr strength, loss and drive tuned). The classical references are cls, the best error over all families and sizes, and eqW, the best error among baselines with at most F_q features, which equalises readout cost. From the curve $q(S)$ we derive: S^*_{eq} and S^*_{cls}, the smallest shot count at which $q(S) \le$ eqW or $\le$ cls; M^*, the smallest ESN whose error is at most q. Wins are declared only if they hold in every seed:

A accuracy win against any classical size, $q <$ cls;

a accuracy win at equal readout cost, $q <$ eqW;

★ scale win: a, $S^*_{\mathrm{eq}} W \le 3 \times 10^5$, and $M^*/F_q \ge 10$ or M^* unreached up to 1000 nodes;

⧫ both: A and $S^*_{\mathrm{cls}} W \le 3 \times 10^5$.

For oracle-input tasks the criteria a, ★ and M^* are not defined, because every stream (quantum and classical) is passed through its own tuned readout whose size is not a property of the estimator, and no ESN family is run on the oracle input; the natural resource is the number of oracle queries per step, and on these tasks "a win on both axes" means the query-matched criterion : (accuracy at matched readout and fewer queries), not ⧫: the quantum estimator at $2m$ queries per step beats every black-box classical estimator with at most $20m$ queries per step, for $m \in \{10^2, 10^3\}$ and every seed. All criteria are recomputed from the raw curves by the report script, so they can be changed post hoc; Fig. 1(a) summarises the rule.

B. Quantum reservoir configurations

All quantum reservoirs are simulated exactly (density matrix, covariance matrix, or truncated Fock space) so that q is free of statistical error; Appendix B lists all 57. Spin reservoirs with $N = 4$–10 qubits use qubit-reset encoding [4], $\rho \to |\psi_u\rangle\langle\psi_u| \otimes \mathrm{Tr}_0 \rho$ with $|\psi_u\rangle = \sqrt{1-u}|0\rangle + \sqrt{u}|1\rangle$, followed by unitary evolution $e^{-iHJ\tau}$ under a free-fermion chain, a random all-to-all XX Hamiltonian, or a strongly non-integrable $XX{+}ZZ$ model with random fields, at coupling $J\tau \in \{0.1, 1, 10\}$ (frozen, edge, scrambled). Readouts are $\langle\sigma^z_i\rangle$, $\langle\sigma^{x,y,z}_i\rangle$, or $\langle\sigma^z_i\rangle$ plus all $\langle\sigma^z_i\sigma^z_j\rangle$, with time multiplexing $V \in \{1, 5, 10\}$. Variants add local T_1/T_2 relaxation as an exact Lindblad propagator, a pseudomode (one extra unmeasured spin coupled by XX to all reservoir spins, $g = 0.5$, damping $T_1 \in \{5, 50\}$) as an engineered non-Markovian environment [25], or an online weak measurement of every readout qubit with strength $g \in \{0.1, 0.3, 0.6\}$, modelled as the ensemble-averaged dephasing channel with coherence factor $\sqrt{1-g^2}$ and signal variance $(1-\langle O\rangle^2)+(1-g^2)/g^2$ per shot [24]. Hamiltonian-encoded spin reservoirs rotate every spin by $\theta = \pi u$ (or $\pi u/4$) about y and rely

on natural T_1/T_2 relaxation for fading memory [7]. The quantum-input reservoir replaces two of its six qubits by the input state at every step.

CV reservoirs comprise an N-mode Gaussian reservoir ($N \in \{4, 12\}$) with squeezed-displaced encoding of mode 0, a fixed random passive interferometer, uniform loss $\eta = 0.8$ and covariance readout [26]; a single driven Kerr mode $H = \chi a^{\dagger 2}a^2 + i\varepsilon(u)(a^\dagger - a)$ with loss, $\chi \in \{0, 0.3, 1\}$; two-mode Fock-space Lindblad reservoirs (8 levels per mode) coupled by cascade or beam splitter, with moment or joint photon-number readout; and quantum Kerr delay lines in the geometry of Ref. [17]: $D \in \{30, 100\}$ non-overlapping pulses circulate through one Kerr element, each pass applying the exact single-mode Lindblad map (16 Fock levels) with Kerr χ, loss κ and drive $\varepsilon_0 u_t m_j$ with a random mask m_j; the ring shift couples the virtual nodes and the readout $\langle x\rangle, \langle p\rangle, \langle n\rangle$ per pulse matches the three features per node of the semiclassical baseline. Two tuned variants choose $(\chi, \kappa, \varepsilon_0)$ per task and seed by the same 20-draw search as the classical baselines [13].

C. Tasks

The 19 classical time-series tasks (1000 training and 500 test steps after a washout of 100; Appendix A) are NARMA2/5/10/20 [27], Mackey–Glass forecasting at horizons 10 and 20 [28], univariate and multivariate Lorenz forecasting [29], linear memory capacity to lag 30 [30], cross-time products $(u_{t-a} - \frac{1}{2})(u_{t-b} - \frac{1}{2})$ [17], delayed XOR, few-data variants of NARMA10 and Mackey–Glass with 100 training steps, and observation-noise variants (1 % and 5 % of the cross-product and XOR tasks, 5 % of NARMA10 and memory capacity). The four quantum-input tasks stream two-qubit states from a collision model in which a fresh Haar-random ancilla partially swaps with the previous ancilla (angle θ_t, piecewise constant over 20 steps) and then with the system; the targets are purity or von Neumann entropy, negativity and θ_t, and the classical baselines receive either the six single-qubit or all fifteen two-qubit Pauli expectations.

The twenty-six oracle-input tasks stream pairs of Boolean functions (f_t, g_t) on $\{0,1\}^n$, and the quantity to track is the Forrelation $\Phi(f, g) = N^{-3/2}\sum_{x,y} f(x)(-1)^{x\cdot y} g(y)$ [31]. Five families are used. In the fresh-pair families a new pair is drawn at every step, so that nothing about it is public or persists between steps: (i) uniformly random $f_t \in \{\pm 1\}^N$ with $g_t = \operatorname{sign} \hat{f}_t$ (n = 10–20, $\Phi_{\mathrm{YES}} = 0.798 \approx \sqrt{2/\pi}$), which is close to but not identical with the Gaussian-rounded distribution of the Aaronson–Ambainis lower bound [31], so that bound is the nearest known result rather than a theorem for this ensemble; (ii) Maiorana–McFarland pairs $f_t(x, y) = (-1)^{x\cdot\pi_t(y)+h_t(y)}$ on $n = 2k$ bits with a fresh random permutation π_t of $\{0,1\}^k$ and random h_t, whose exact dual gives $\Phi_{\mathrm{YES}} = 1$ (n = 10–16), masked by a fresh uniformly random invertible $A_t \in \mathrm{GL}_n(\mathbb{F}_2)$ applied jointly, $(f_t \circ A_t,\ g_t \circ A_t^{-\mathsf{T}})$, which leaves Φ unchanged and is the masking step of the hardness construction of Ref. [32]; and (iii) the same Maiorana–McFarland pairs without the mask, whose fixed coordinate split is exposed to a classical querier (Sec. III.E), kept as a structured control. The two fixed families keep one pair per task: (iv) Kloosterman (n = 10–20) and (v) Gold pairs (n = 10–16). Kloosterman pairs take $f = (-1)^{\mathrm{Tr}(x^{-1})}$ over $\mathrm{GF}(2^n)$ and $g = \operatorname{sign} \hat{f}$, giving $\Phi_{\mathrm{YES}} \approx 0.85$ and a phase that is not quadratic; Gold pairs take the bent function $f = (-1)^{\mathrm{Tr}(s_3 x^3)}$ and $g = \operatorname{sign} \hat{f}$, giving $\Phi_{\mathrm{YES}} = 1$ with both functions quadratic over GF(2) [33, 34]. A hidden two-state Markov chain (switch probability 0.1) alternates between YES pairs, scrambled at every step by a random element of the invariance group $(\chi_c f,\ g \circ \tau_c)$ that leaves Φ unchanged, and NO pairs in which g is additionally translated by a random $e \neq 0$ so that $\Phi \approx 0$; the targets are Φ_t, the hidden state, and a NARMA10 series driven by $(1 + \Phi_t)/8$. The two drift-monitoring tasks replace the Markov chain by a device g_t whose fidelity $p_t \in [0.3, 1]$ to the ideal dual $g = \operatorname{sign} \hat{f}$ follows an Ornstein–Uhlenbeck walk with jumps, so that $\mathbb{E}[\Phi_t] = p_t \Phi_{\mathrm{YES}}$; the targets are p_t and Φ_t. Exact Φ_t for every translation follows from two fast Walsh–Hadamard transforms (Appendix D). Classical-input and quantum-input cells have three seeds, except 32 key cells (the tuned $D = 100$ Kerr delay line on four tasks, the ten-qubit reservoir on two, and 26 of the 39 equal-cost wins found at three seeds) which have ten; oracle-input cells have ten. The known-f and white-box estimators of the next subsection apply only to the fixed families; in the fresh-pair families there is no description to use, but the unmasked Maiorana–McFarland structure admits an exact adaptive attack.

D. Oracle-input estimators

The quantum feature map is the fixed two-query circuit $H^{\otimes n} D_g H^{\otimes n} D_f H^{\otimes n}$ followed by a Hadamard test, whose acceptance probability is exactly $(1 + \Phi)/2$ [31]; the feature at step t is the acceptance frequency over m shots, each shot costing two oracle queries, with binomial noise and $W = 1$. Black-box classical estimators with Q queries per step are (i) naive sampling, $\hat{\Phi} = \sqrt{N}\, f(x)(-1)^{x\cdot y} g(y)$ over Q random pairs, unbiased with variance N/Q; (ii) the all-pairs U-statistic, which queries f at $Q/2$ points x_i and g at $Q/2$ points y_j and averages $f(x_i)(-1)^{x_i\cdot y_j} g(y_j)$ over all $Q^2/4$ pairs (the signs come from one sparse Walsh–Hadamard transform), an implemented $\sqrt{N}$-query estimator with variance $4/Q + 4N/Q^2$ that saturates the query lower bound $\Omega(\sqrt{N}/\log N)$ for constant-advantage estimation up to logarithmic factors [31, 35]; and (iii) a degree-2 algebraic-normal-form (ANF) attack that learns the ANF of f_t and g_t from $1 + n + n(n-1)/2$ queries each, verifies it on 32 random points and, if both verify, computes Φ_t exactly, otherwise falling back to naive sampling. Two

white-box estimators use the printed definition of the oracles: (iv) for the Forrelation tasks the known-f estimator computes $\hat{f}$ once per task (one Walsh–Hadamard transform, no queries), recovers c_t from $n+1$ queries of f_t at 0 and the basis vectors, and spends the remaining $K = Q-(n+1)$ queries on $g_t(z \oplus c_t)$ at points z drawn with probability $\propto |\hat{f}(z)|$, returning $A\,\overline{\operatorname{sign}\hat{f}(z)}\, g_t(z \oplus c_t)$ with $A = N^{-3/2}\sum_z |\hat{f}(z)|$, which is unbiased for Φ_t with variance at most A^2/K independent of N; (v) for the drift tasks a monitor that knows the ideal g samples $g_t(y)g(y)$ at Q points, with variance $(1-p_t^2)/Q$; and (vi) for the unmasked Maiorana–McFarland pairs an exact adaptive black-box attack: $k+1$ queries of f_t at $(0,0)$ and $(u_j,0)$ give $\pi_t(0)$, $k+1$ queries of g_t at $(\pi_t(0),0)$ and $(\pi_t(0),u_j)$ give $\pi_t^{-1}(\pi_t(0)\oplus\alpha)$, which is nonzero exactly when the translation has $\alpha \neq 0$, and otherwise $k+1$ queries of f_t per basis vector u_j give $\pi_t(u_j)$ and one query of g_t at $(\pi_t(u_j),0)$ gives β_j; the pair is a YES instance exactly when $\alpha=\beta=0$, which for a bent pair determines $\Phi_t \in \{0,1\}$ exactly with at most k^2+4k+2 queries; on the masked pairs the same queries see no coordinate blocks and the attack is at chance. The quantum frequency stream and every classical estimate stream pass through the same tuned readout family (ESN, poly-NARX and Kerr delay line at a reduced size grid, 10 draws), so that the comparison isolates the feature-extraction stage; the full-information classical baseline that receives exact Φ_t as input uses the same reduced readout grid and enters cls.

III. RESULTS

A. Classical-input map: no win on either axis

Figure 1 shows the map. Over the 989 (configuration, task) pairs with classical input (51 configurations on the 19 classical time-series tasks and 5 quantum-input configurations on the 4 quantum-input tasks; three seeds each, ten for 32 key cells), no cell satisfies A, ★ or ⧫. Over all pairs the quantum error is a median of 2.6 decades above the best classical error, with an interquartile range of 1.2 to 5.1 decades; for the best quantum configuration on each task the median gap is 0.81 decades (interquartile range 0.29–1.79). The best classical model is the semi-classical Kerr delay line in 1422 of the 3191 classical-input cells, the ESN in 951, and the poly-NARX in 818; the poly-NARX solves the noiseless XOR and cross-product tasks exactly (NMSE $\sim 10^{-24}$) because the targets are low-degree polynomials of lagged inputs. Representative best-quantum-per-task errors are 6.3×10^{-5} against 2.8×10^{-9} for Mackey–Glass at horizon 10, 0.035 against 0.013 for NARMA10, 0.131 against 0.100 for NARMA20 and a tie of 0.0229 against 0.0230 for 5 %-noise XOR.

Equal readout cost changes the picture only marginally. 26 pairs satisfy a, with error ratios eqW/q between 1.02 and 26 (the largest for the Hamiltonian-encoded chain on multivariate Lorenz, 5.1×10^{-4} against 1.35×10^{-2} at 90 features). Only 8 of the 26 reach the equal-cost baseline within 10^4 shots, 4 more within 10^6, and 14 only with exact expectations; 13 of the 26 have $q>0.5$ and 11 have a margin below 10 %; the 13 equal-cost wins seen at three seeds that did not hold at ten are one four-qubit and three six-qubit z-readout reservoirs on memory capacity and its noisy variant, the single Kerr mode on the three cross-product tasks, and the online weak zz readout on Mackey–Glass 20 and noisy XOR; and in every one of them an ESN of at most 100 nodes (median over seeds; 300 in the worst seed) reaches the quantum error, so none meets the $M^*/F_q \geq 10$ clause of ★. Observation noise removes the exact classical solutions but not the ranking: with 5 % noise the best quantum errors are 0.076 against 0.059 for cross products, 0.074 against 0.055 for NARMA10 and 0.058 against 0.039 for memory capacity.

The axes of the map rank the design elements. Measured as the median over the 19 classical time-series tasks of $\log_{10} q$/cls for the six-qubit random-coupling reservoir with z readout, the coupling strength (frozen 3.00, edge 2.99, scrambled 2.92), the Hamiltonian class (chain 3.00, random 2.99, chaotic 3.00) and the qubit number ($N=4,6,8$: 2.99, 2.99, 2.99) each change the median by less than 0.1 decade. Two-body zz readout (2.72), time multiplexing ($V=5$: 2.75, $V=10$: 2.61, and 2.19 with zz), and natural relaxation in Hamiltonian encoding (strong T_1/T_2 2.07, weak 2.56, none 3.01) improve accuracy by 0.3–0.9 decade, and the 12-mode Gaussian reservoir with covariance readout is the most accurate quantum family on classical input (1.06 decades), although its covariance propagation is classically efficient. The pseudomode improves short-memory nonlinear tasks (Mackey–Glass 10: z readout 0.104 to 0.080, zz readout 0.0127 to 0.0062) and degrades long linear memory (memory capacity 0.858 to 0.872, NARMA20 0.524 to 0.594). Increasing to $N=10$ with zz readout (55 features) lowers the Mackey–Glass error from 0.0127 to 1.2×10^{-3} (ten seeds), which is above the equal-cost classical error of 9.2×10^{-4}, is matched by an ESN of 30 to 100 nodes depending on the seed, and is not reached at any finite shot count up to 10^6. In the two-mode Fock-space reservoirs, Kerr cascade (0.41–0.45) and photon-number readout (0.60 against 0.86 at $\chi=0$) both help, but the equal-cost classical error is 0.019 and the ten-node semi-classical Kerr delay line reaches 0.0063.

B. Shot wall: readout family, not reservoir, sets the cost

Figure 2 plots the finite-shot error against the exact one for every pair. The degradation at the budget is heavy-tailed and family-dependent. For the quadrature readout of CV Kerr reservoirs the median ratio $q(10^4)/q(\infty)$ is 1.02 with a 90th percentile of 1.3 and a maximum of 3.1 over 342 pairs; for the one-bit Forre-

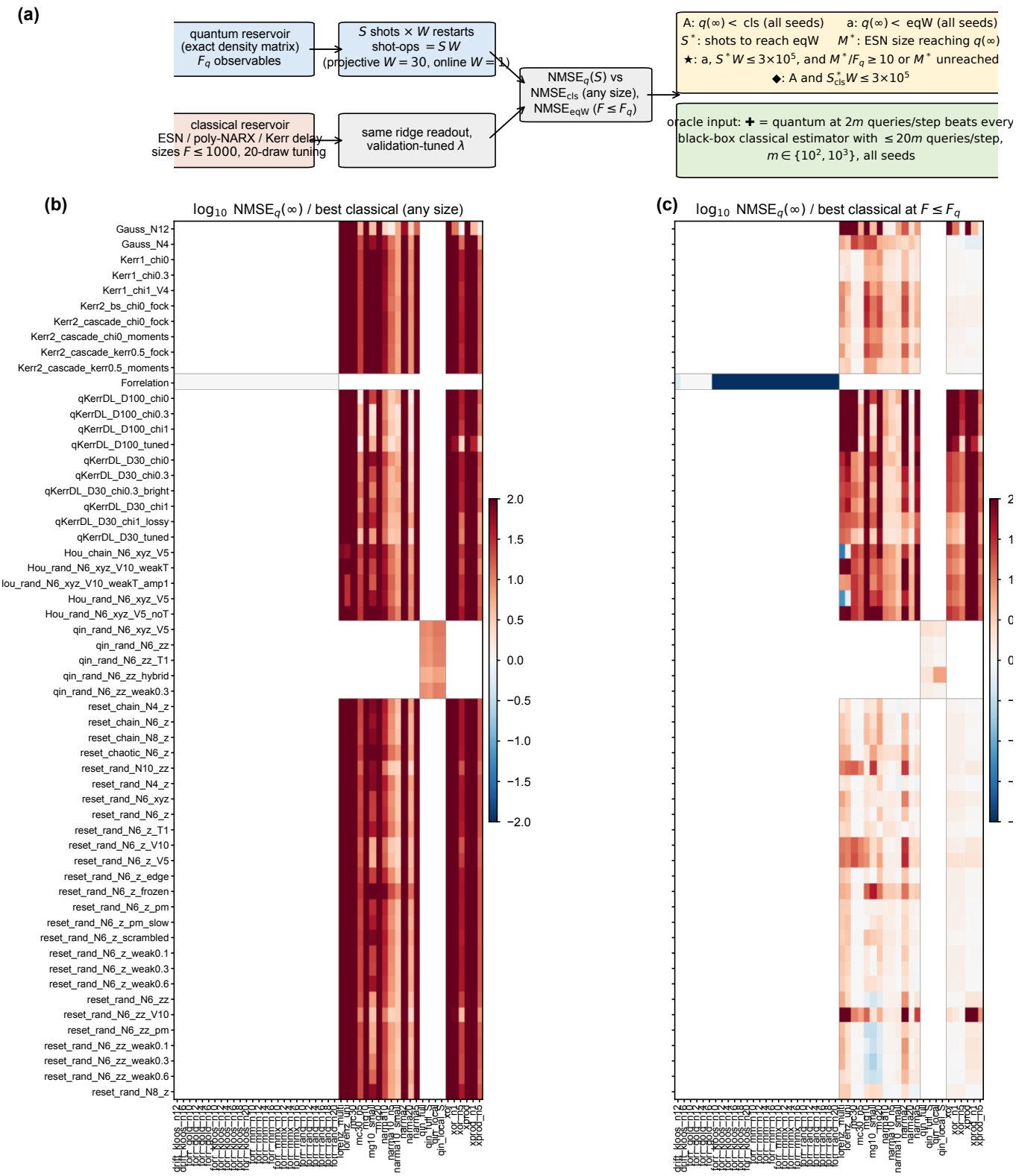


FIG. 1. **No classical-input or quantum-input cell beats the tuned classical baselines, and no cell meets the scale criterion.** (a) Accounting rule: quantum and classical reservoirs share the ridge readout; the quantum side pays S shots per feature and W restarts per step; wins A, a, ★, ♦ and : are defined in Methods. (b) Median over seeds of $\log_{10} q(\infty)/\text{cls}$ for all 57 configurations (rows) and 49 tasks (columns; three seeds per cell, ten for the oracle tasks and 32 key classical-input cells); red means the quantum reservoir is less accurate than the best classical model of any size. White cells are incompatible (quantum-input reservoirs run only on quantum-input tasks, the Forrelation circuit only on oracle tasks). (c) The same against the best classical model with at most F_q features; no cell satisfies ★ (oracle cells are excluded because a, eqW and M^* are undefined there, Methods). Colour scale clipped at ±2 decades.

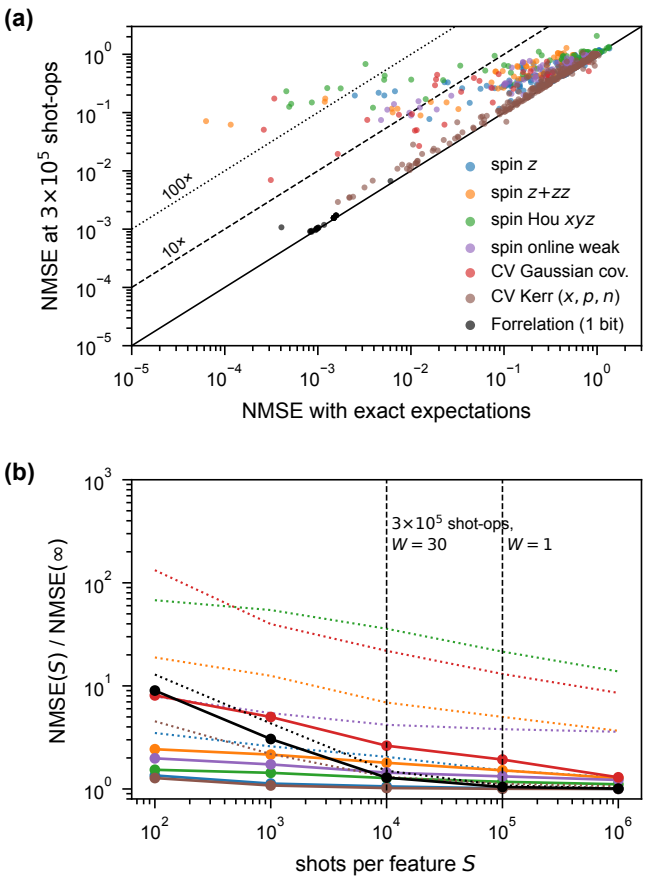


FIG. 2. **At the 3×10^5 shot-operation budget, the most accurate spin and covariance readouts lose up to three orders of magnitude where quadrature and one-bit readouts lose about a factor of three at most.** (a) Error at the budget ($S = 10^4$ for projective, $S = 10^5$ for online readouts) against the exact-expectation error for every (configuration, task) pair, coloured by readout family; the solid line is no degradation, dashed and dotted lines mark 10- and 100-fold degradation. (b) Median (solid) and 90th percentile (dotted) of $q(S)/q(\infty)$ per family; vertical lines mark the budget for $W = 30$ and $W = 1$.

lation readout the median at $S = 10^5$ is 1.05. For spin z readouts the median is 1.06 but the 90th percentile is 2.1 and the maximum 85; for zz readouts 1.79, 6.9 and 900; for Hamiltonian-encoded xyz readouts 1.28, 36 and 507; for Gaussian covariance readouts 2.63, 21.9 and 435; for the online weak readouts at $S = 10^5$ 1.33, 3.8 and 23. The tail is where the accuracy is: among cells with $q(\infty) < 10^{-2}$, the median degradation is 129 for spin zz (7 pairs, range 16–1133), 43 for spin z (4 pairs), 191 for Hamiltonian-encoded (10 pairs), 22 for Gaussian covariance (6 pairs), 18 for online weak (4 pairs), 1.6 for CV Kerr (11 pairs, range 1.06–1.8) and 1.1 for the Forrelation readout (26 pairs). The reason is the ratio of signal variation to single-shot variance: a Pauli expectation has a single-shot variance of up to one while its useful variation is of order 10^{-2} at the error levels that compete with classical models, whereas $\langle x\rangle$, $\langle p\rangle$ and $\langle n\rangle$ of a driven mode vary by an amount comparable to their single-shot spread. Counting replays, an equal-cost spin win at $S^* = 10^5$ costs $STW = 10^5 \times 1600 \times 30 \approx 5 \times 10^9$ shot-operations for the 1600-step task, compared with about 10^4 multiply–adds per step for the sparse 1000-node ESN of Appendix C (ten connections per node). These are modelled costs under the independent-noise, uncharged-settings convention of Methods. Charging measurement settings and multiplexing runs as well (Appendix E) lowers the shots available at the budget to $S = 333$–10^4 for the spin and CV families and leaves the ordering unchanged: the accurate-cell degradation becomes 59 for spin z, 129 for spin zz, 315 for Hamiltonian-encoded xyz, 30 for Gaussian covariance, 2.3 for CV Kerr and 1.0 for the one-bit readout.

Online weak measurement removes the replay factor but not the wall. At equal shot-operations ($S = 10^5$, $W = 1$ against $S = 10^4$, $W = 30$) the six-qubit zz reservoir with $g = 0.3$ reaches 0.47 against 0.435 on NARMA10, 0.099 against 0.089 on Mackey–Glass 10 and 0.112 against 0.116 on Mackey–Glass 20 ($g = 0.6$: 0.089 and 0.0745); $g = 0.1$ is worse (0.157 on Mackey–Glass 10) because the added pointer variance dominates. At $S = 10^6$ the back-action keeps the weak readout at 0.0745 where the projective one reaches 0.0469. The online variants add equal-cost wins but each is matched by an ESN of 10 to 65 nodes at the median over seeds.

C. Kerr delay line: quantum beats its mean field, not its tuned twin

The Kerr delay line is the one configuration whose classical counterpart is the same equations with a coherent amplitude in place of the density matrix, so it allows a three-way comparison (Fig. 3). At identical parameters and $\chi > 0$ the exact quantum model is more accurate than the mean-field twin on all ten tasks (medians over ten seeds): the ratio of mean-field to quantum error is 2.8 for $D = 30$, $\chi = 0.3$ (range over tasks 1.2–24), 2.9 for $\chi = 1$ (1.2–44), 2.4 for the bright variant with $\chi = 0.1$, $\varepsilon_0 = 1.5$ (1.4–42) and 3.9 for $D = 100$, $\chi = 0.3$ (1.7–65). Examples are NARMA10 at 0.266 against 0.806 and Mackey–Glass 10 at 0.017 against 0.41 for $D = 30$, and 3.0×10^{-3} against 0.19 for $D = 100$. At $\chi = 0$ the dynamics is linear, the mean field is exact for the moments, and the ratio is 0.9 (range 0.38–1.0), i.e. the quantum model is equal or worse; at $\chi = 0$ the undriven-loss dynamics lets the photon number reach a median of 4.5 and a maximum of 12.4, so the 16-level Fock truncation, not readout noise, degrades the quantum features. The peak mean photon number in these runs has a median over seeds of 1.2–1.6 (maximum 2.4) for $\varepsilon_0 = 0.5$ and $\chi > 0$, and about 9 (maximum 14) in the bright variant, so the gain is a few-photon effect: Kerr phase diffusion and non-Gaussian states generate features that a coherent amplitude cannot.

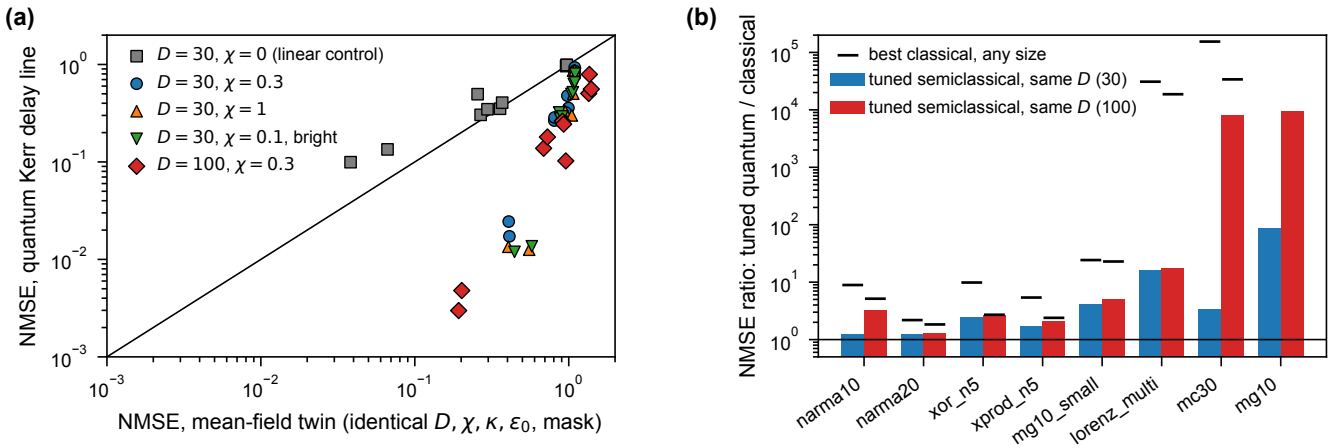


FIG. 3. **Few-photon Kerr dynamics beats the mean-field model with identical parameters but loses to the semiclassical model once the classical parameters are tuned.** (a) Exact quantum Kerr delay line against its mean-field twin (same D, χ, κ, ε_0 and mask, same three features per pulse, same readout) on ten tasks; points below the diagonal favour the quantum model. At $\chi = 0$ the two coincide up to the Fock-space truncation of the quantum model. (b) Error of the quantum Kerr delay line with $(\chi, \kappa, \varepsilon_0)$ tuned by the 20-draw search, divided by the error of the semiclassical discrete delay-line baseline of the same D tuned with the same budget (bars) and by the best classical model of any size (black ticks); medians over ten seeds. Data from `kerr_twin.json` and `tune_qkerr.json`.

The gain does not survive tuning of the classical side, except as a tie on nonlinear-memory tasks. The semiclassical delay-line baseline is a discrete map (Appendix C), not the mean-field twin itself, so we compare against both. When $(\chi, \kappa, \varepsilon_0)$ of the quantum model are chosen by the same 20-draw search as the semiclassical baseline of the same D, the tuned quantum error exceeds the tuned semiclassical one at the median on every task and depth: by factors 1.3 and 3.2 on NARMA10 ($D = 30$, 100), 1.2 and 1.3 on NARMA20, 2.4 and 2.7 on 5 %-noise XOR, 1.7 and 2.1 on 5 %-noise cross products, 4.1 and 5.0 on few-data Mackey–Glass, 16 and 17 on multivariate Lorenz, 3.3 and 7.9×10^3 on memory capacity, and 87 and 9.4×10^3 on Mackey–Glass 10 (medians over ten seeds). The seed-to-seed spread is large on the smooth tasks: the interquartile range of the tuned quantum error is 12–143 % of its median and that of the tuned semiclassical error 5–1547 %, so the two largest ratios are uncertain by an order of magnitude; the ordering nevertheless holds in 8–10 of 10 seeds on every task (the quantum model is better in 1 of 10 seeds on NARMA10 at $D = 30$ and multivariate Lorenz at $D = 100$, and in 2 of 10 on NARMA20). Tuning the mean-field twin itself with the identical protocol gives the same verdict: on the smooth deterministic tasks the tuned twin is 3 to 2300 times more accurate than the tuned quantum model (Mackey–Glass 10: 57 and 2300 for $D = 30$ and 100; memory capacity 4.3 and 15; multivariate Lorenz 4.9 and 2.8; few-data Mackey–Glass 6.7 and 7.2; medians over ten seeds, quantum better in 0–3 of 10), whereas on NARMA10, NARMA20 and the 5 %-noise XOR and cross-product tasks the two are within a factor 0.8–1.4 of each other at the median and the quantum model is better in 1 to 5 of 10 seeds. The search lets the classical amplitudes avoid the parameter region in which the mean field fails, and on smooth deterministic targets the classical model retains continuous amplitudes without quantum noise and gains two to four decades. The tuned $D = 100$ quantum delay line is nevertheless the most accurate quantum configuration on memory capacity (0.026 at ten seeds), but never below the semiclassical delay line with the same 300 features.

D. Quantum-input tasks: polynomial readouts of expectations win

The quantum-input tasks were designed to disadvantage classical models, which see only Pauli expectations of each input state. They did not. A poly-NARX of degree 2 on the six single-qubit expectations reaches a mean error of 0.091 over purity, negativity and θ_t, and 0.092 when purity is replaced by von Neumann entropy; with all fifteen expectations the best classical error is 0.113 and 0.126 (1000-node ESN). The quantum reservoir with linear readout is linear in each input state and so cannot represent purity or entropy exactly; it can only exploit the temporal correlations of the stream, and its mean error is 0.94 with zz readout, 1.05 with xyz readout and multiplexing, 0.89 with relaxation and 0.95 with online weak measurement. A hybrid in which a fixed 100-node ESN post-processes the quantum features [36] lowers the error to 0.60, still 6.5 times the classical value. Purity is a quadratic polynomial of expectations, and entropy, though not a polynomial, is smooth enough to be fitted

by one; the restricted classical access is a weaker constraint than the linearity of the quantum readout.

E. Oracle-input tasks: the only wins on both axes

Figure 4 and Table I give the oracle-input results (ten seeds per cell). In the fresh-pair family the quantum estimator reaches an error of 1.07–1.14 $\times 10^{-2}$ at 200 queries per step and 4.2–4.9 $\times 10^{-3}$ at 2000 for every n, independent of N, while the all-pairs classical estimator reaches 2.2×10^{-3}, 2.1×10^{-3}, 2.8×10^{-3}, 5.2×10^{-3}, 7.1×10^{-3} and 1.7×10^{-2} at $Q = 20000$ and 7.2×10^{-3}, 1.0×10^{-2}, 0.025, 0.13, 0.40 and 0.68 at $Q = 2000$ for $n = 10$ to 20; naive sampling stays near or above 0.1 and the ANF attack fails (error about 1) at every n. Against every black-box estimator with at most ten times its queries the quantum side at $m = 10^2$ is better in 0, 2, 10, 10, 10 and 10 of 10 seeds for $n = 10$ to 20, and at $m = 10^3$ in 0, 1, 1, 10, 10 and 10 of 10 seeds, so the strict criterion : holds for $n = 16$, 18 and 20 ($N \geq 6.5\times10^4$), with a median margin at $m = 10^3$ of 1.2, 1.7 and 4.0 and at $m = 10^2$ of 12, 40 and 60. Here no classical estimator other than the black-box ones exists, because the pair has no description and does not persist, and the per-step query lower bound of Ref. [31] applies. The interpolated crossover grows with N with a fitted exponent of 0.47 over n = 10–18. The masked Maiorana–McFarland family behaves in the same way with a different constant: the quantum error is 2.6×10^{-3} at 200 queries and 1.7×10^{-3} at 2000 at every n (its readout floor is 1.5×10^{-3} because $\Phi_{\rm YES} = 1$ sets a different target scale, so floors are not comparable across families), the all-pairs estimator reaches 1.8, 2.0, 2.6 and 4.4×10^{-3} at $Q = 20000$ and 5.3×10^{-3}, 7.4×10^{-3}, 0.018 and 0.071 at $Q = 2000$ for $n = 10$ to 16, naive sampling stays above 0.05, and the ANF attack fails (0.91–1.03). The quantum side is better in 10 of 10 seeds at $m = 10^2$ for every n and in 8, 8, 10 and 10 of 10 seeds at $m = 10^3$, so : holds at $n = 14$ and 16; the adaptive attack of Methods misclassifies about half of the steps on this family (median 0.48) and its readout error is 1.0. The interpolated crossover is 9.1×10^3 and 1.2×10^4 at $n = 10$ and 12 and beyond the grid above, too few points for an exponent.

The unmasked Maiorana–McFarland family shows what the mask protects. Against the sampling estimators it behaves like the masked one: the all-pairs estimator reaches 1.7, 1.9, 2.8 and 5.0 $\times 10^{-3}$ at $Q = 20000$ and the ANF attack fails (0.89–1.03), and the quantum side is better in 10 of 10 seeds at $m = 10^2$ for every n and in 4, 8, 9 and 10 of 10 seeds at $m = 10^3$, a nominal : at $n = 16$. The family nevertheless has a classical shortcut: the adaptive attack of Methods recovers Φ_t exactly at every step with at most 47, 62, 79 and 98 queries for $n = 10$ to 16 (mean 29–57), fewer than the 200 queries of the quantum estimator at $m = 10^2$, and its error through the same readout is the full-information floor 1.5×10^{-3}, below the quantum estimator at every finite shot count. The unmasked family therefore joins the Gold and Kloosterman families as a structured input with a classical shortcut, and the two fresh families without exposed structure, uniformly random pairs and masked Maiorana–McFarland pairs, are the only ones on the map whose wins survive every implemented classical estimator. Against the sampling estimators the crossover budget Q^* at which the classical median first reaches the quantum error at 200 queries grows with N in the random, masked Maiorana–McFarland and Kloosterman families (Fig. 4f): interpolated on the grid it is 1.5×10^3, 1.9×10^3, 4.7×10^3, 1.1×10^4 and 1.6×10^4 for random pairs at n = 10–18, 9.1×10^3 and 1.2×10^4 for masked Maiorana–McFarland pairs at $n = 10$ and 12, and beyond the 2×10^4 grid at the larger n.

The fixed families show what the black-box restriction hides. For Kloosterman pairs the quantum estimator reaches an error of 9.4–10.0 $\times 10^{-3}$ at 200 queries per step and 3.6–3.9 $\times 10^{-3}$ at 2000 for every n from 10 to 20, because its variance $(1-\Phi^2)/m$ does not depend on N. The description-blind classical estimators depend on N: the all-pairs estimator reaches 0.19, 6.7×10^{-3} and 1.7×10^{-3} at $Q = 200$, 2000 and 20000 for $n = 10$, but 1.0, 0.67 and 0.014 for $n = 20$; naive sampling is at 0.08 for $n = 10$ and 1.0 for $n = 20$ at $Q = 20000$. The classical query count Q^* at which the all-pairs median first drops below the quantum error at 200 queries is 2000 for $n = 10$ and 12, 20000 for n = 14–18 and above 20000 for $n = 20$; the count $Q_{\rm fit}$ at which the two estimator variances are equal, the solution of $4/Q+4N/Q^2 = (1-\Phi^2)/m$, which approaches $2\sqrt{mN/(1-\Phi^2)}$ for large N, is 2124, 3242, 5614, 10433, 20109 and 39482 for $n = 10$ to 20 with $\Phi = 0.849$, i.e. 11 to 197 times the quantum budget; the grid values of Q^* agree with it to within the decade spacing of the grid (the first grid value reaching the quantum error is 2000, below $Q_{\rm fit}$, at $n = 10$ and 12; 20000, above $Q_{\rm fit}$ at $n = 14$ and 16 and just below it at $n = 18$; and beyond the grid at $n = 20$), and the interpolated crossover grows with a fitted exponent of 0.48. Against these estimators the quantum side wins at the median at both budgets only for $n \geq 16$: at $m = 10^2$ (200 queries) against every description-blind estimator with at most 2000 queries it is better in 0, 2, 10, 10, 10 and 10 of 10 seeds for $n = 10$ to 20, and at $m = 10^3$ against estimators with at most 20000 queries in 0, 0, 0, 9, 10 and 10 of 10 seeds. The strict query-matched criterion : therefore holds only for $n = 18$ and 20; at $n \leq 14$ the all-pairs estimator at 20000 queries is more accurate than the quantum estimator at 2000 in every seed (while at 200 against 2000 queries the quantum side already wins every seed from $n = 14$), and at $n = 16$ the quantum margin at $m = 10^3$ is 1.4-fold and reverses in one seed. Criterion A is not met, because the full-information baseline that receives exact Φ_t through the same readout grid gives exactly the infinite-shot quantum error (1.0×10^{-3}), as it must.

On the fixed family the black-box restriction is the whole of the advantage. The known-f estimator, which

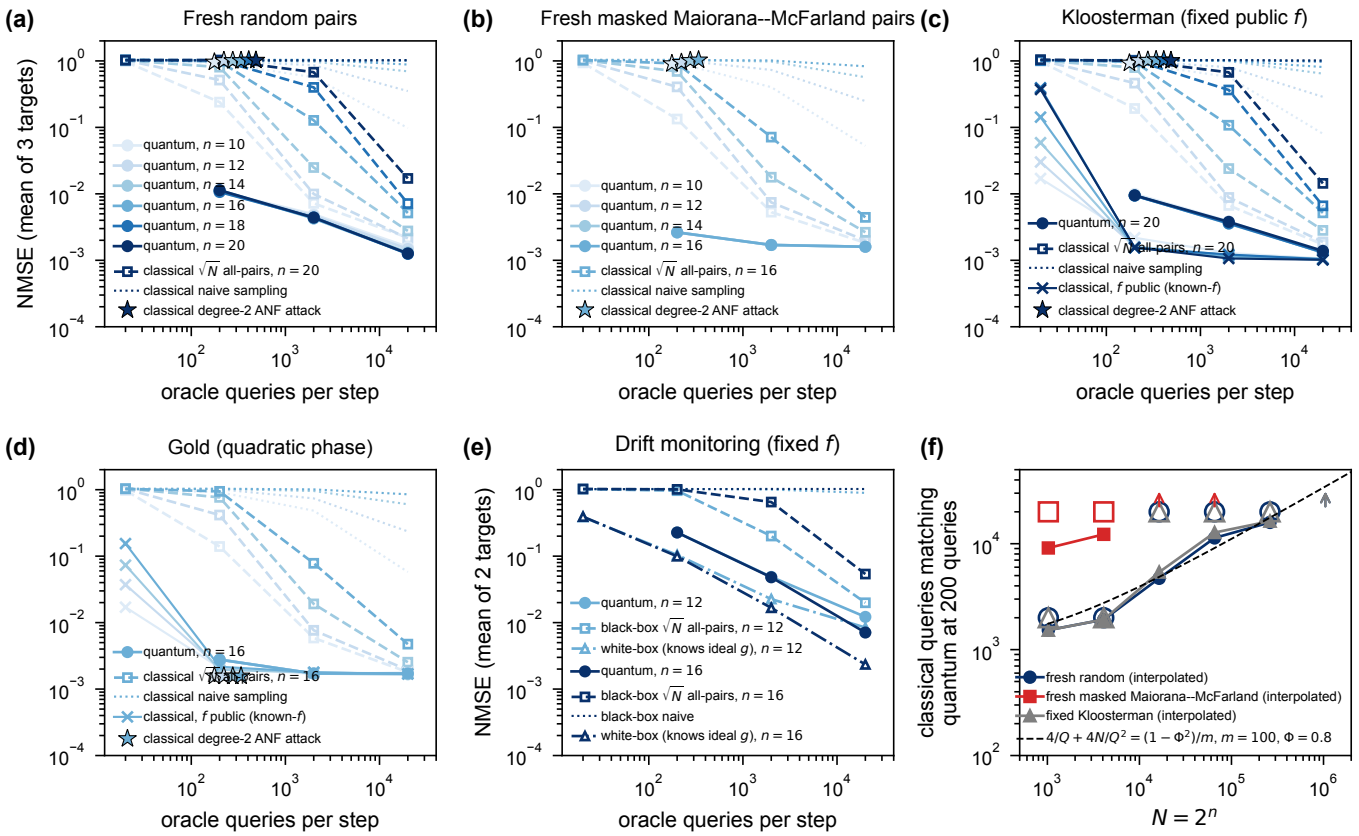


FIG. 4. **On fresh random oracle pairs the quantum Forrelation estimator wins from $N = 2^{16}$ and the matching classical query budget grows as $\sqrt{N}$; on a fixed public pair a classical estimator that uses f beats it at equal queries at every N, quadratic phases fall to a degree-2 ANF attack, and a monitor that knows the ideal relation beats the quantum drift monitor.** Error of the tracked targets against oracle queries per step, medians over ten seeds, same tuned readout on every stream. (a) Fresh random pairs, $n = 10$–20: quantum circuit at $2m$ queries ($m = 10^2, 10^3, 10^4$ shots), classical naive sampling, the all-pairs $\sqrt{N}$-query estimator and the degree-2 ANF attack (stars). (b) Fresh affinely masked Maiorana–McFarland pairs, $n = 10$–16, same estimators; the unmasked family (Table I) is not plotted because its sampling-estimator curves coincide with these and the exact adaptive attack reaches the full-information floor at every n. (c) Fixed Kloosterman pairs, $n = 10$–20: quantum circuit at $2m$ queries ($m = 10^2, 10^3, 10^4$ shots; independent of N), classical naive sampling, the all-pairs $\sqrt{N}$-query estimator (its crossing with the quantum curve moves right as $\sqrt{N}$), the known-f estimator (crosses) and the degree-2 ANF attack at its own query count (stars, which fail and fall back to naive sampling). (d) Gold pairs: the ANF attack recovers Φ_t exactly and matches the quantum circuit at infinite shots. (e) Drift monitoring of a device against a Kloosterman reference, $n = 12$ and 16: black-box classical estimators, the quantum monitor, and the white-box estimator that samples $g_t g$. (f) Classical query budget at which the all-pairs median first reaches the quantum error at 200 queries, against N: filled symbols interpolated on the grid, open symbols the first grid value, arrows beyond the grid (masked Maiorana–McFarland: two points); the dashed line is the variance-matching prediction, the solution of $4/Q + 4N/Q^2 = (1 - \Phi^2)/m$ with $m = 100$ and $\Phi = 0.8$.

uses the printed definition of f and no other information, reaches an error of 2.2×10^{-3} at 200 queries for $n = 10$ and 1.6×10^{-3} for $n = 12$, and 1.5–1.6×10^{-3} for $n = 14$–20, below the quantum estimator at 200 queries in every seed at every n and below the quantum estimator at 2000 queries in 9 or 10 of 10 seeds. Its error does not grow with N, because on YES steps $g_t(z \oplus c_t) = \operatorname{sign}\hat{f}(z)$ makes the estimate exact and on NO steps its variance is at most A^2/K. The quantum circuit is identical on all three families and at all N; what the fixed families test is therefore not a separation between quantum and classical computation but the cost of classical estimation without the description of f, and because f is a fixed, publicly specified function there, the white-box hardness assumed by the black-box model is false, not merely unproven. The fresh-pair family removes the description and the persistence, and only there does the win survive every implemented classical estimator.

For Gold pairs the win disappears. Both functions are quadratic over GF(2), so the ANF attack reconstructs them exactly from 176–338 queries and computes Φ_t without error; its error of 1.6×10^{-3} equals that of the quantum circuit at infinite shots and is below the quantum error at $m = 10^4$ (1.7×10^{-3}). The same attack

TABLE I. Oracle-input tasks ($N = 2^n$, ten seeds; Fresh = uniformly random pairs, MMx = affinely masked Maiorana–McFarland pairs, MM = the same pairs unmasked, all fresh at every step; for MM the attack column gives the exact adaptive attack of Methods, which reaches the full-information floor with at most 98 queries, so its : is nominal (parenthesised): error (median) of the quantum estimator at m shots per step ($2m$ queries), of the all-pairs description-blind classical estimator at Q queries per step, and of the known-f estimator at $Q = 200$ (inapplicable, –, to fresh pairs; for MM the adaptive attack; for the drift tasks: the white-box monitor at $Q = 200$); ANF = degree-2 attack at its own query count (176, 222, 276, 338, 408, 486 queries for $n = 10$–20). : = query-matched win over the description-blind estimators in every seed. Sources: `results/forr_matched.md`, `results/forr_knownf.json`.

task	m=10^2	m=10^3	Q=200	Q=$2{\cdot}10^4$	known-f / ANF	
Fresh n=10	0.0109	0.0049	0.24	0.0022	– / 0.96	
Fresh n=12	0.0107	0.0047	0.52	0.0021	– / 1.0	
Fresh n=14	0.0110	0.0044	0.81	0.0028	– / 1.0	
Fresh n=16	0.0114	0.0042	0.98	0.0052	– / 1.0	:
Fresh n=18	0.0107	0.0044	1.0	0.0071	– / 1.0	:
Fresh n=20	0.0113	0.0044	1.0	0.017	– / 1.0	:
MMx n=10	0.0026	0.0017	0.13	0.0018	– / 0.91	
MMx n=12	0.0026	0.0017	0.41	0.0020	– / 0.95	
MMx n=14	0.0026	0.0017	0.69	0.0026	– / 1.0	:
MMx n=16	0.0026	0.0017	0.91	0.0044	– / 1.0	:
MM n=10	0.0026	0.0017	0.14	0.0017	0.0015 / 0.89	
MM n=12	0.0026	0.0017	0.41	0.0019	0.0015 / 0.99	
MM n=14	0.0026	0.0017	0.71	0.0028	0.0015 / 1.0	
MM n=16	0.0026	0.0017	0.89	0.0050	0.0015 / 1.0	(:)
Kloos. n=10	0.0100	0.0039	0.19	0.0017	0.0022 / 0.93	
Kloos. n=12	0.0094	0.0038	0.46	0.0019	0.0016 / 1.0	
Kloos. n=14	0.0095	0.0036	0.79	0.0028	0.0016 / 1.0	
Kloos. n=16	0.0094	0.0037	0.97	0.0052	0.0015 / 1.0	
Kloos. n=18	0.0094	0.0036	1.0	0.0066	0.0015 / 1.0	:
Kloos. n=20	0.0095	0.0038	1.0	0.014	0.0016 / 1.0	:
Gold n=10	0.0028	0.0017	0.14	0.0018	0.0021 / 0.0016	
Gold n=16	0.0028	0.0017	0.94	0.0047	0.0022 / 0.0016	
Drift n=12	0.23	0.049	0.97	0.020	WB 0.10	
Drift n=16	0.23	0.048	1.0	0.054	WB 0.10	

fails verification on Kloosterman pairs (errors 0.93–1.04) and degenerates to naive sampling. The non-quadratic phase denies the classical side one shortcut, the ANF reconstruction, but not the other: the known-f estimator is as effective on Kloosterman pairs as on Gold pairs.

The drift-monitoring tasks locate the win in the knowledge of the classical side. In the black-box setting the quantum monitor reaches 0.23 at 200 queries and 0.048 at 2000 for $n = 16$, against 1.0 and 0.65 for the all-pairs estimator and 0.054 at 20000 queries; at $n = 16$ the quantum side wins at the median at both budgets (10 of 10 seeds at $m = 10^2$, 6 of 10 at $m = 10^3$, where the 20000-query classical estimator is within 1.1-fold), and at $n = 12$ it loses at the median at both budgets (3 of 10 and 0 of 10 seeds), so the strict : criterion is not met on either drift task. The white-box estimator that knows the ideal g reaches 0.10 at 200 queries, 2.3 times better than the quantum monitor at equal queries (2.9 and 3.0 times at 2000 and 20000; 1.5–2.2 at $n = 12$), because sampling $g_t(y)g(y)$ has variance $(1 - p_t^2)/Q$ with no dependence on N, while each quantum shot spends two queries. The split between the black-box median win at $n = 16$ and the white-box loss is the same statement as for the Forrelation tasks: the advantage exists only when the ideal relation is unavailable to the classical monitor.

IV. DISCUSSION

A. One condition

Across 3451 cells the quantum reservoirs that win on both axes share one property: the input is a fresh instance with no structure exposed to the classical querier, for which no classical shortcut is known. Every classical-input and quantum-input task lost, whatever the reservoir. On the oracle-input tasks the target is a functional with a proven query separation (Forrelation, $O(1)$ quantum against $\Omega(\sqrt{N}/\log N)$ classical queries [31, 35]); in the fresh random family, the nearest simulable analogue of that bound's distribution, the quantum estimator wins in every seed from $N = 2^{16}$ against every implemented classical estimator and the matching query count grows with N with a fitted exponent of 0.47; in the masked Maiorana–McFarland family, the masking construction of Ref. [32], it wins in every seed from $N = 2^{14}$. The bound covers none of the structured families: a classical estimator that reads the printed definition of a fixed Kloosterman f matches or beats the quantum estimator at equal queries at every N, a purely black-box learner can read the whole table of a fixed f within the stream for $n \leq 18$, and the unmasked Maiorana–McFarland pairs are solved exactly by an adaptive attack within 98 queries, which the affine mask defeats. Freshness and an inaccessible description are therefore not sufficient; the instance must also expose no structure along which to query. Non-quadratic phases defeat one classical shortcut (degree-2 reconstruction) and not the other; the one-bit readout of known variance is what keeps the quantum shot cost polynomial, but the same property is shared by any Pauli measurement. No reservoir property is needed for the win: the coupling-strength, Hamiltonian-class and qubit-number axes move the median accuracy by less than 0.1 decade (readout, multiplexing and relaxation by 0.3–0.9 decade without closing the gap), and the memoryless Forrelation circuit, which has no dynamics at all, is the only winner. The win is also a scaling statement rather than a fixed-N one: against the sampling estimators with ten times the quantum query budget the quantum side wins in every seed at both budgets only from $N = 2^{14}$ (masked Maiorana–McFarland pairs), 2^{16} (random pairs) or 2^{18} (fixed Kloosterman pairs); below that it wins at the 200-query budget first (from $n = 14$ for random and Kloosterman pairs, from $n = 10$ for Maiorana–McFarland pairs) and still loses at the 2000-query budget; and for every family with exposed structure a different, exact classical

algorithm removes the win altogether.

These findings sharpen, and partly contradict, the necessary conditions distilled from the literature. The requirement that the reservoir be classically non-simulable at low readout order [18, 19] is not met by any reservoir on the map; the requirement of polynomial shot cost [23, 24] is met by the one-bit readout; the tuned-baseline protocol of Ref. [13] is built into every cell, with the asymmetry (classical side tuned per size, quantum configurations fixed) running in the classical side's favour. The Kerr result (Sec. III.C) shows what a single-model comparison measures: the quantum model beats the classical limit of the same equations at the same parameters, in agreement with the separation of Kerr feedback from Gaussian linear optics [17], but a classical model with freely tuned parameters is a stronger baseline than the untuned classical limit: the tuned discrete delay line beats the tuned quantum model at the median on every task, and the tuned twin itself does so on the smooth tasks and ties on the nonlinear-memory tasks.

B. Application classes

Two application classes were examined and both are conditional on the information asymmetry. The first is black-box process verification: two processes f and g_t, accessible only through queries, should satisfy a Fourier-dual relation, and the fidelity of g_t to that relation must be tracked over time. The drift-monitoring tasks are a prototype; the quantum monitor wins at the median at $N = 2^{16}$ against classical monitors that can neither compute nor sample the ideal relation, and loses by a factor 2–3 at equal queries to one that can, so the application requires processes that are genuinely black boxes, as in the fresh-pair family, not fixed public functions. The second is the evaluation of large-field exponential-sum correlations. The exact classical computation of Φ for a Kloosterman pair by two Walsh–Hadamard transforms takes 0.65 s at $n = 24$ on one core and scales as $\log_2(\text{time}) \approx 1.22n - 30$ over the measured range, while the quantum estimator needs 622 queries (636 at $n = 10$) for a precision of ± 0.03 at every n; however, the known-f estimator reaches the same precision with a comparable number of queries once $\hat{f}$ is available, and the YES value itself converges to 0.8488 for $n \geq 18$, so the classical cost is that of one transform per field, not per instance, and the application survives only for oracles that are physical processes without a known description. Two further candidates were examined and rejected: code synchronisation, because the receiver knows the reference code and the white-box estimator wins, and Boolean matching under affine equivalence, because the invariance group of Φ acts on the pair jointly and does not define a one-sided similarity.

C. Limitations

Four limitations bound the claims. The oracle-input wins are established in the query model only: query counts are the resource on both sides and the classical references are implemented estimators (naive sampling and the all-pairs U-statistic with variance $4/Q + 4N/Q^2$). In the fresh random family the Aaronson–Ambainis bound is the nearest known result, but it concerns the decision problem on a Gaussian-rounded distribution, not the Boolean sign-Walsh ensemble simulated here, so the random-family win is a statement about the implemented estimators, and a better-constant estimator would move Q^* without removing its growth. For the fixed families the exclusion of classical preprocessing is not justified by any hardness result: f is a fixed public function, the general Forrelation lower bound [31, 32] does not transfer to this structured family, and the known-f estimator shows that the family has no white-box hardness. The black-box model is also undermined by amortisation: because f is fixed across the stream, a black-box learner can spend its per-step budget on reading the table of f itself, and with $Q = 200$ queries per step and 1600 steps it holds the full table ($3.2 \times 10^5 \geq N$) for every $n \leq 18$ before the stream ends, after which it runs the known-f estimator; the rigorous setting for a query separation is a fresh random pair per step, the Aaronson–Ambainis distribution, in which neither the known-f estimator nor amortisation exists and the $\Omega(\sqrt{N}/\log N)$ per-step bound applies. Second, the oracle implementation cost is excluded: a Kloosterman phase needs a $\mathrm{GF}(2^n)$ inversion circuit per query and the dual $g = \operatorname{sign}\hat{f}$ needs a coherent oracle of its own, which no physical process is known to provide; for the fresh-pair families the provider of the dual oracle must itself hold $\hat{f}_t$, so the advantage is over a classical estimator that queries the provider, not over the provider. Third, the shot model charges S shots per feature with independent noise and a fixed replay of 30 steps, without charging measurement settings or multiplexing runs; it under-charges the quantum side, so the negative conclusions are conservative but the per-family shot-wall ranking is a modelled, not a measured, protocol cost. Fourth, on the classical-input side the simulation method (exact density matrices and single-mode Fock spaces) bounds the reservoirs to $N \leq 10$ spins and to delay lines whose pulses do not overlap, the classical-input cells have three seeds except the 32 key cells, and the classical baselines are tuned per size while all but two quantum configurations are fixed; multi-mode Kerr feedback with overlapping pulses and $N \geq 20$ spins with correlated readouts lie outside the map, and whether they win is the question of their classical simulability itself.

V. EXPERIMENTAL FEASIBILITY

The map is a simulation, but each of its components maps onto a platform on which the protocol can be run,

and the accounting rule fixes the experimental cost. Table II converts the dimensionless parameters into physical units for the reference frequencies stated there; the numbers are estimates from those reference values, not measurements (script `scripts/feasibility_units.py`, output `paper/feasibility_units.json`).

Mapping. The qubit-reset spin reservoirs correspond to superconducting qubit arrays with mid-circuit reset and projective readout [8, 11], or to nuclear spins with ensemble readout [7]; the Hamiltonian-encoded reservoirs are the rotation encoding of the NMR experiment. The quantum Kerr delay line is a single Kerr-nonlinear resonator with a delayed feedback path, the single-node delay-line geometry of classical reservoir computing [2], in the few-photon regime that a superconducting Kerr resonator reaches. The Forrelation estimator is n qubits, three Hadamard layers and two calls of a phase oracle per shot; the YES/NO decision needs only the probability of the all-zero outcome, which is Φ^2 (0.64 for random pairs, 1 for Maiorana–McFarland pairs), so no control qubit or controlled oracle is required unless the sign of Φ matters. The online weak readouts correspond to dispersive measurement at reduced strength.

Reachability. With $J/2\pi = 10$ MHz a reservoir step is 16 ns and the natural decay per step at $T_1 = 100$ μs is 1.6×10^{-4}, so the no-relaxation configurations of the map are the native regime of superconducting qubits, and the configurations with relaxation in 3–5 steps require engineered dissipation (a tunable coupling to a lossy mode) rather than natural decay. The Kerr delay line at $K/2\pi = 1$ MHz needs a round trip of 48 ns, a linewidth of 0.17 MHz ($Q \approx 3\times10^4$ at 5 GHz) and a drive of 1.7 MHz, all within the range of three-dimensional superconducting cavities with a weakly coupled nonlinear element; the demanding element is the 4.8 μs delay of the $D = 100$ loop, which in a microwave circuit is a storage mode or a sequence of pulses rather than a transmission line. The same model at few photons is not reachable in optical fibre, where the Kerr phase per pass is many orders of magnitude below $\chi = 0.3$. The Forrelation estimator at $n = 16$ uses 16 qubits and 200 oracle calls per step; its requirement is not the circuit but the oracle, a physical process that applies a phase to a superposition of inputs, which is what distinguishes an oracle input from a classical time series.

Cost. At the 3×10^5 shot-operation budget a projective-readout spin reservoir costs 0.3 s per step and 8 min per 1600-step task with 1 μs readout-and-reset cycles, and the 10^6-shot column of the map costs 13 h per task; with 1 s cycles, the scale of nuclear-spin readout, the same budget is 3.5 days per step, so the finite-shot columns of the map are a superconducting or trapped-ion experiment and the ensemble-readout experiments correspond to the exact-expectation column.

Observables. Three predictions of the map are testable with existing readouts. (i) The shot wall: on one chip holding a Kerr resonator and a qubit array, the most accurate Kerr quadrature cells degrade by a median factor 1.6 at the budget while the most accurate two-body spin readouts degrade by 129. (ii) The mean-field gap: at a mean photon number of 1.2–1.6 the Kerr delay line should beat its mean-field model, with fixed parameters, by a median factor 2.8–3.9, and the gap should close as the Kerr strength is tuned to zero. (iii) The access-model prediction: with fresh random oracle pairs the quantum estimator's error at 100 shots per step should be independent of n while the classical crossover budget grows with N, measurable on 10–16 qubits.

Other platforms. Trapped ions with cycles of order 10 μs multiply the shot costs above by ten; neutral-atom arrays [9] with millisecond cycles by 10^3; photonic reservoirs [10] have fast detection but their covariance readout is the most shot-sensitive family on the map.

VI. CONCLUSION

We have built the first map of quantum reservoir computing that places 57 exactly simulated reservoir configurations, 49 tasks, six shot levels and three classical families tuned by random search under one accounting rule that counts accuracy, readout cost, shots and restarts together. Across the 989 classical-input and quantum-input pairs no quantum reservoir beats the best classical model, by a median of 2.6 decades over pairs and 0.8 decades for the best configuration per task, and the 26 equal-cost wins are each matched by an echo-state network of at most 100 nodes at the median over seeds; the few-photon Kerr delay line beats its own mean field at identical parameters by up to 65-fold, yet once the twin is tuned it loses by up to three orders of magnitude on smooth tasks and only ties on nonlinear-memory tasks; and the Forrelation cells with fresh oracle pairs that expose no structure, uniformly random pairs and affinely masked Maiorana–McFarland pairs, win on both axes against every implemented classical estimator in every seed from $N = 2^{16}$ and 2^{14}, with the matching classical query budget growing with N (as $\sqrt{N}$ for random pairs over the simulated range), whereas every family with exposed structure (fixed Kloosterman, Gold and unmasked Maiorana–McFarland pairs) falls to a family-specific classical algorithm with fewer queries than the quantum circuit. Compared with single-model reports of advantage [7, 10, 17], the map reproduces the effects of relaxation as fading memory, two-body readout, Kerr nonlinearity and non-Gaussian readout as accuracy gains of 0.3–0.9 decade against other quantum configurations, and shows that none of them closes the gap to a tuned classical model. The result places quantum reservoir advantage in the access model of the input rather than in dynamical complexity: on this map it appears only for fresh instances with no exposed structure, and every input with exposed structure admitted a classical shortcut. If this is ignored, the field will continue to tune reservoirs toward operating points that move the median by a tenth of a decade while leaving the three decades of shot

TABLE II. Dimensionless map parameters in physical units at the stated reference values (estimates from `paper/feasibility_units.json`; $T = 1600$ steps per task).

quantity	map value	physical estimate
spin step time, $J\tau = 1$	–	16 ns at $J/2\pi = 10$ MHz; 1.6 ms at $J/2\pi = 100$ Hz
natural decay per step	0 (no-relaxation configs)	1.6×10^{-4} at $T_1 = 100$ μs, 16 ns steps
engineered $T_1 = 5$ steps	0.2 per step	80 ns at 16 ns steps
budget 3×10^5 shot-ops	$S = 10^4$, $W = 30$	0.3 s per step, 8 min per task at 1 μs cycles
$S = 10^6$, $W = 30$	3×10^7 shot-ops	13 h per task at 1 μs cycles
Kerr round trip, $\chi = 0.3$	–	48 ns at $K/2\pi = 1$ MHz
Kerr loss, $\kappa = 0.05$ per pass	–	$\kappa/2\pi = 0.17$ MHz, $Q = 3 \times 10^4$ at 5 GHz
Kerr drive, $\varepsilon_0 = 0.5$	$\langle n \rangle \approx 1.2$–$1.6$	$\varepsilon_0/2\pi = 1.7$ MHz
delay-line memory, $D = 100$	100 pulses	4.8 μs loop
Forrelation, $n = 16$	$m = 100$ shots	16 qubits, 200 oracle calls per step

cost in place, and will continue to report black-box wins on fixed instances that a description-aware classical estimator erases; if it is adopted, the design effort shifts to inputs that are fresh black-box instances, physical processes without a learnable description, and to reservoirs beyond exact simulation. The next steps are a lower bound for the Boolean sign-Walsh ensemble actually simulated, masked Maiorana–McFarland streams at $n \geq 18$, an accounting of the oracle circuit cost on both sides, and exact simulation of multi-mode Kerr feedback with overlapping pulses, the one reservoir class this map could not reach.

ACKNOWLEDGMENTS

Data availability. The data that support the findings of this article (the full map of 3451 cells with its finite-shot curves, the tuned classical baselines, and the Kerr-twin, Kerr-tuning, tuned-mean-field, Forrelation, known-f and tuning-sensitivity result files behind Figs. 1–4 and Tables I–III) and the source code that generated them (the `qrcfit` package with its map driver, report, Kerr-twin, Kerr-tuning, mean-field-tuning, Forrelation-report, known-f, application-demo and figure scripts) are publicly available from Ref. [37], a single Zenodo deposit that also holds this manuscript (all versions: DOI 10.5281/zenodo.23589289); every number quoted in this article is listed with its source file in the `numbers.json` file produced by the figure script.

Appendix A: Supplemental material: tasks

Table III lists every task with its dimensions (three seeds per cell; ten for the Forrelation-family tasks), the best quantum configuration and the best classical baseline. NARMAn follows Ref. [27] with input $u \in [0, 0.5]$ (tanh-bounded for $n = 20$); Mackey–Glass uses $\tau = 17$, $dt = 0.1$, subsampling 10; Lorenz uses $dt = 0.01$ with subsampling 2 and horizon 10; memory capacity sums the squared correlations up to lag 30; cross products use lag pairs $(1, 2), (1, 3), (2, 5), (3, 8)$; delayed XOR uses delays $1, 3, 5, 10$; noise variants add Gaussian noise of 1 % or 5 % of the input range to the input with clean targets; few-data variants use 100 training steps. The quantum-input stream uses system–ancilla swap angle $\phi = \pi/4$ and ancilla–ancilla angle $\theta_t \in [0, \pi/2]$ redrawn every 20 steps. The oracle streams are described in Methods; the fresh-pair streams draw f_t uniformly from $\{\pm 1\}^N$ (random family) or as a Maiorana–McFarland function with a fresh permutation and h_t at every step, with or without a fresh random affine mask $A_t \in \mathrm{GL}_n(\mathbb{F}_2)$ applied jointly to the pair, with the same hidden Markov YES/NO switching as the fixed families. The 32 classical-input cells with ten seeds are the tuned $D = 100$ Kerr delay line on memory capacity, its noisy variant, noisy XOR and NARMA20; the ten-qubit zz reservoir on Mackey–Glass 10 and NARMA10; and 26 of the 39 equal-cost wins found at three seeds (including the two Hamiltonian-encoded reservoirs on multivariate Lorenz), of which 13 held at ten seeds; the remaining 13 three-seed wins keep three seeds.

Appendix B: Supplemental material: quantum configurations

Table IV lists the 57 configurations. Spin reservoirs use $J\tau = 1$ unless marked frozen (0.1) or scrambled (10), field $h = 1$, and a fresh random Hamiltonian per seed; relaxation uses $T_1 = T_2 = 5$ (reset encoding) or $T_1 = 3, T_2 = 2$ (strong) and $T_1 = 30, T_2 = 10$ (weak) in units of the step time (Hamiltonian encoding). Gaussian reservoirs use squeezing $r = x - \frac{1}{2}$ and displacement $d = 2x$ for input x. Single-mode Kerr reservoirs use $\kappa = 0.3$, $\varepsilon_0 = 1.5$, 20 Fock

TABLE III. All 49 tasks (generated by `scripts/paper_figures.py` from `qrcfit/tasks.py`, `results/map.jsonl` and `results/baselines.json`; medians over seeds: three for classical-input and quantum-input tasks, ten for oracle tasks and for 32 key classical-input cells).

task	d_{in}	d_{out}	T_{train}	T_{test}	best quantum (F_q)	NMSE_q	best classical (NMSE)
drift_kloos_n12	1	2	1000	500	forr_qrc (1)	0.00604	forr_full:4096 (0.00604)
drift_kloos_n16	1	2	1000	500	forr_qrc (1)	0.000406	forr_full:65536 (0.000406)
forr_gold_n10	1	3	1000	500	forr_qrc (1)	0.00157	forr_full:1024 (0.00157)
forr_gold_n12	1	3	1000	500	forr_qrc (1)	0.00157	forr_full:4096 (0.00157)
forr_gold_n14	1	3	1000	500	forr_qrc (1)	0.00157	forr_full:16384 (0.00157)
forr_gold_n16	1	3	1000	500	forr_qrc (1)	0.00157	forr_full:65536 (0.00157)
forr_kloos_n10	1	3	1000	500	forr_qrc (1)	0.00102	forr_full:1024 (0.00102)
forr_kloos_n12	1	3	1000	500	forr_qrc (1)	0.000999	forr_full:4096 (0.000999)
forr_kloos_n14	1	3	1000	500	forr_qrc (1)	0.000945	forr_full:16384 (0.000945)
forr_kloos_n16	1	3	1000	500	forr_qrc (1)	0.001	forr_full:65536 (0.001)
forr_kloos_n18	1	3	1000	500	forr_qrc (1)	0.000973	forr_full:262144 (0.000973)
forr_kloos_n20	1	3	1000	500	forr_qrc (1)	0.000961	forr_full:1048576 (0.000961)
forr_mm_n10	1	3	1000	500	forr_qrc (1)	0.00148	forr_full:1024 (0.00148)
forr_mm_n12	1	3	1000	500	forr_qrc (1)	0.00148	forr_full:4096 (0.00148)
forr_mm_n14	1	3	1000	500	forr_qrc (1)	0.00148	forr_full:16384 (0.00148)
forr_mm_n16	1	3	1000	500	forr_qrc (1)	0.00148	forr_full:65536 (0.00148)
forr_mmx_n10	1	3	1000	500	forr_qrc (1)	0.00148	forr_full:1024 (0.00148)
forr_mmx_n12	1	3	1000	500	forr_qrc (1)	0.00148	forr_full:4096 (0.00148)
forr_mmx_n14	1	3	1000	500	forr_qrc (1)	0.00148	forr_full:16384 (0.00148)
forr_mmx_n16	1	3	1000	500	forr_qrc (1)	0.00148	forr_full:65536 (0.00148)
forr_rand_n10	1	3	1000	500	forr_qrc (1)	0.00117	forr_full:1024 (0.00117)
forr_rand_n12	1	3	1000	500	forr_qrc (1)	0.000964	forr_full:4096 (0.000964)
forr_rand_n14	1	3	1000	500	forr_qrc (1)	0.000829	forr_full:16384 (0.000829)
forr_rand_n16	1	3	1000	500	forr_qrc (1)	0.000861	forr_full:65536 (0.000861)
forr_rand_n18	1	3	1000	500	forr_qrc (1)	0.000858	forr_full:262144 (0.000858)
forr_rand_n20	1	3	1000	500	forr_qrc (1)	0.00083	forr_full:1048576 (0.00083)
lorenz_multi	3	1	1000	500	spin_hamenc_chain_N6_xyz_V5 (90)	0.000517	esn:1000 (7.28e-06)
lorenz_uni	1	1	1000	500	spin_hamenc_rand_N6_xyz_V5 (90)	0.00516	esn:300 (0.000172)
mc30	1	30	1000	500	qkerr_delay_D100_tuned (300)	0.026	kerr_delay:300 (7.66e-07)
mc30_n5	1	30	1000	500	qkerr_delay_D100_chi0 (300)	0.0579	kerr_delay:100 (0.0389)
mg10	1	1	1000	500	spin_reset_rand_N6_zz_V10 (210)	6.29e-05	kerr_delay:300 (2.77e-09)
mg10_small	1	1	100	300	qkerr_delay_D100_chi0.3 (300)	0.00905	kerr_delay:300 (0.00471)
mg20	1	1	1000	500	spin_reset_rand_N6_zz_V10 (210)	0.000116	esn:1000 (9.57e-08)
narma10	1	1	1000	500	cv_gauss_N12 (324)	0.0346	kerr_delay:300 (0.0134)
narma10_n5	1	1	1000	500	cv_gauss_N12 (324)	0.0742	kerr_delay:100 (0.0554)
narma10_small	1	1	100	300	qkerr_delay_D100_chi0.3 (300)	0.281	polynarx:(30, 1) (0.143)
narma2	1	1	1000	500	cv_gauss_N12 (324)	0.000312	esn:1000 (5.73e-06)
narma20	1	1	1000	500	qkerr_delay_D100_chi0 (300)	0.131	kerr_delay:300 (0.1)
narma5	1	1	1000	500	cv_gauss_N12 (324)	0.00164	kerr_delay:300 (0.000111)
qin_full	15	3	1000	500	spin_qin_rand_N6_zz_hybrid (221)	0.595	esn:1000 (0.113)
qin_full_S	15	3	1000	500	spin_qin_rand_N6_zz_hybrid (221)	0.596	esn:1000 (0.126)
qin_local	6	3	1000	500	spin_qin_rand_N6_zz_hybrid (221)	0.595	polynarx:(3, 2) (0.0912)
qin_local_S	6	3	1000	500	spin_qin_rand_N6_zz_hybrid (221)	0.596	polynarx:(3, 2) (0.0922)
xor	1	4	1000	500	cv_gauss_N12 (324)	0.011	polynarx:(20, 2) (4.38e-25)
xor_n1	1	4	1000	500	cv_gauss_N12 (324)	0.0115	kerr_delay:100 (0.00101)
xor_n5	1	4	1000	500	cv_gauss_N12 (324)	0.0229	kerr_delay:100 (0.023)
xprod	1	4	1000	500	cv_gauss_N12 (324)	0.00743	polynarx:(10, 2) (6.6e-24)
xprod_n1	1	4	1000	500	cv_gauss_N12 (324)	0.0103	polynarx:(10, 2) (0.00248)
xprod_n5	1	4	1000	500	cv_gauss_N12 (324)	0.0757	polynarx:(10, 2) (0.0592)

levels; two-mode reservoirs use $\chi \in \{0, 0.5\}$, $\kappa = 0.3$, coupling $g = 0.3$, 8 levels per mode; quantum Kerr delay lines use $\kappa = 0.05$, $\varepsilon_0 = 0.5$ (bright: $\chi = 0.1$, $\varepsilon_0 = 1.5$, 24 levels; lossy: $\kappa = 0.3$), with the drive quantised on a 33-level grid of precomputed propagators. The tuned delay lines search $\chi \in [10^{-1.5}, 10^{0.5}]$, $\kappa \in [10^{-1.7}, 10^{-0.3}]$, $\varepsilon_0 \in [10^{-0.7}, 10^{0.3}]$ (log-uniform, 20 draws, mean photon number capped at 9 for truncation safety).

Appendix C: Supplemental material: classical baselines

ESN: sparse random recurrent matrix (10 connections per node), spectral radius $\rho \in [0.3, 1.5]$, input scale $\in [10^{-1}, 10^{0.5}]$, leak $\in [0.1, 1]$, tanh nodes, sizes $3, 5, 10, 30, 100, 300, 1000$. Poly-NARX: monomials up to degree p of L lagged inputs, sizes $(L, p) \in \{(2,1),(3,1),(5,1),(10,1),(30,1),(3,2),(5,2),(10,2),(20,2),(5,3),(10,3)\}$, skipped when the feature count exceeds 5000. Semiclassical Kerr delay line: $D \in \{10, 30, 100, 300\}$ virtual nodes with random mask, Kerr $\chi \in [10^{-1}, 10]$, loss $\eta \in [0.5, 0.99]$, drive $\varepsilon \in [10^{-1}, 10^{0.5}]$, features $|a_j|^2, \mathrm{Re}\, a_j, \mathrm{Im}\, a_j$. Each (family, size, task, seed) receives 20 random draws; the configuration with the best validation error is scored once on the test segment. Oracle-task readouts use the reduced grid ESN $\{10, 30, 100, 300\}$, poly-NARX $\{(5,1),(10,2),(20,2),(30,1)\}$, Kerr delay $\{10, 30, 100\}$ with 10 draws, applied identically to quantum and classical streams. Classical black-box estimators use $Q \in \{20, 200, 2000, 20000\}$ queries per step. Tuning-budget sensitivity: raising the ESN random search from 20 to 50 draws (best error over all sizes, three seeds) leaves NARMA10 and 5 %-noise XOR unchanged (ratios 1.00 and 1.05) and strengthens the classical side on Mackey–Glass 10 (0.16), memory capacity (0.32), multivariate Lorenz (0.07) and NARMA20 (0.57); no task changes the sign of the quantum–classical comparison, so the 20-draw map is conservative in the classical side's disfavour.

Appendix D: Supplemental material: $\mathrm{GF}(2^n)$ instances and the degree-2 ANF attack

Field arithmetic uses exponential and logarithm tables of $\mathrm{GF}(2^n)$ built from a primitive polynomial for each n (e.g. $x^{16} + x^{15} + x^{13} + x^4 + 1$ for $n = 16$); the absolute trace $\mathrm{Tr}(a) = \sum_{i<n} a^{2^i}$ is tabulated from the same tables, and $f^K(x) = (-1)^{\mathrm{Tr}(x^{-1})}$ with $f^K(0) = 1$. The Gold function uses $f(x) = (-1)^{\mathrm{Tr}(s_3 x^3)}$ with n even and s_3 a non-cubic residue, which makes f bent and $\Phi(f, \mathrm{sign}\,\hat{f}) = 1$. For either family $g = \mathrm{sign}\,\hat{f}$ where $\hat{f}$ is the Walsh–Hadamard transform. The correlations of f with every translate of g follow from $S_e = \widehat{(f \cdot \hat{g})}(e)$, two transforms of length N, and $\Phi_e = S_e / N^{3/2}$; the invariance $\Phi(\chi_c f, g \circ \tau_c) = \Phi(f, g)$ is used to scramble each YES step. The ANF attack queries 0, e_i and $e_i \oplus e_j$ $(1 + n + n(n-1)/2$ points) and recovers the degree-≤ 2 coefficients by Möbius inversion restricted to degree 2: $a_0 = b(0)$, $a_i = b(e_i) \oplus a_0$, $A_{ij} = b(e_i \oplus e_j) \oplus a_0 \oplus a_i \oplus a_j$ with $b = (1 - f)/2$; the reconstructed table is accepted if it agrees with the oracle on 32 further random points, so that the total query count per step is $2(1 + n + n(n-1)/2 + 32)$, i.e. 176, 222, 276 and 338 for $n = 10, 12, 14, 16$. For a quadratic f the reconstruction is exact; for the Kloosterman phase, whose algebraic degree is $n - 1$, the verification fails with probability close to one.

Appendix E: Supplemental material: weak-measurement and shot model

Finite-shot features are $X + \xi\sqrt{V/S}$ with ξ standard normal, drawn independently for every feature, time step and shot level and applied to training and test segments alike; V is the exact single-shot variance $\langle O^2 \rangle - \langle O \rangle^2$ (Pauli: $1 - \langle O \rangle^2$; covariance features: the Gaussian fourth-moment expression). The online weak measurement of strength g is modelled by its ensemble-averaged channel, the dephasing with Kraus operators $\sqrt{(1 \pm c)/2}\,\{I, \sigma^z\}$, $c = \sqrt{1 - g^2}$, applied to every readout qubit after every sub-step, and by an added pointer variance $(1 - g^2)/g^2$ per shot; conditional trajectories are not simulated. The budget of 3×10^5 shot-operations corresponds to $S = 10^4$ with $W = 30$ replays or $S = 10^5$ with $W = 1$. Assumptions of the model: (i) every feature is estimated from its own S shots with independent noise, although z and zz features share the same z-basis shots and xyz readouts need three settings; (ii) time-multiplexed projective readouts ($V > 1$) would need V separate runs per step, which are not charged; (iii) the replay length $W = 30$ is a convention and is not checked against the fading memory of each configuration (the quantum Kerr delay line with $\kappa = 0.05$ retains $e^{-1.5} \approx 0.22$ of its amplitude after 30 passes). The first two under-charge the quantum side; the third can err in either direction depending on the configuration's memory. The two Hamiltonian-encoded configurations with weak relaxation (`weakT`) are projective readouts with $W = 30$. As a check, the figure script also evaluates a settings-charged budget, shot-ops $= S\,W\,s\,V$ with $s = 3$ settings for xyz and for $\langle x \rangle, \langle p \rangle, \langle n \rangle$ readouts, $s = 2$ for Gaussian covariance readouts and $s = 1$ otherwise, and V separate runs for time multiplexing; at 3×10^5 shot-ops this leaves $S = 333$ (Hamiltonian-encoded, $V = 10$) to 10^4 shots, interpolated on the simulated shot grid. The family medians of $q(S)/q(\infty)$ (same statistic as in Sec. III.B) move from 1.06/1.79/1.28/2.63/1.02 to

TABLE IV. All 57 quantum configurations (generated by `scripts/paper_figures.py`).

config	readout family	F_q	W	med. $\log_{10} q$/cls
cv_gauss_N12	CV Gaussian cov.	324	30	1.06
cv_gauss_N4	CV Gaussian cov.	44	30	2.47
cv_kerr_chi0	CV Kerr (x,p,n)	4	30	2.97
cv_kerr_chi0.3	CV Kerr (x,p,n)	4	30	2.97
cv_kerr_chi1_V4	CV Kerr (x,p,n)	16	30	2.84
cv2_bs_chi0_fock	CV Kerr (x,p,n)	16	30	2.92
cv2_cascade_chi0_fock	CV Kerr (x,p,n)	16	30	2.95
cv2_cascade_chi0_moments	CV Kerr (x,p,n)	8	30	2.97
cv2_cascade_kerr0.5_fock	CV Kerr (x,p,n)	16	30	2.87
cv2_cascade_kerr0.5_moments	CV Kerr (x,p,n)	8	30	2.91
forr_qrc	Forrelation (1 bit)	1	1	0.00
qkerr_delay_D100_chi0	CV Kerr (x,p,n)	300	30	2.93
qkerr_delay_D100_chi0.3	CV Kerr (x,p,n)	300	30	2.89
qkerr_delay_D100_chi1	CV Kerr (x,p,n)	300	30	2.76
qkerr_delay_D100_tuned	CV Kerr (x,p,n)	300	30	1.76
qkerr_delay_D30_chi0	CV Kerr (x,p,n)	90	30	2.97
qkerr_delay_D30_chi0.3	CV Kerr (x,p,n)	90	30	2.97
qkerr_delay_D30_chi0.3_bright	CV Kerr (x,p,n)	90	30	2.90
qkerr_delay_D30_chi1	CV Kerr (x,p,n)	90	30	2.91
qkerr_delay_D30_chi1_lossy	CV Kerr (x,p,n)	90	30	2.42
qkerr_delay_D30_tuned	CV Kerr (x,p,n)	90	30	2.31
spin_hamenc_chain_N6_xyz_V5	spin Hou xyz	90	30	1.85
spin_hamenc_rand_N6_xyz_V10_weakT	spin Hou xyz	180	30	2.56
spin_hamenc_rand_N6_xyz_V10_weakT_amp1	spin Hou xyz	180	30	2.50
spin_hamenc_rand_N6_xyz_V5	spin Hou xyz	90	30	2.07
spin_hamenc_rand_N6_xyz_V5_noT	spin Hou xyz	90	30	3.01
spin_qin_rand_N6_xyz_V5	spin z	90	30	1.01
spin_qin_rand_N6_zz	spin z+zz	21	30	0.96
spin_qin_rand_N6_zz_T1	spin z+zz	21	30	0.94
spin_qin_rand_N6_zz_hybrid	spin z+zz	221	30	0.77
spin_qin_rand_N6_zz_weak0.3	spin online weak	21	1	0.97
spin_reset_chain_N4_z	spin z	4	30	3.00
spin_reset_chain_N6_z	spin z	6	30	3.00
spin_reset_chain_N8_z	spin z	8	30	3.00
spin_reset_chaotic_N6_z	spin z	6	30	3.00
spin_reset_rand_N10_zz	spin z+zz	55	30	2.61
spin_reset_rand_N4_z	spin z	4	30	2.99
spin_reset_rand_N6_xyz	spin z	18	30	3.00
spin_reset_rand_N6_z	spin z	6	30	2.99
spin_reset_rand_N6_z_T1	spin z	6	30	2.94
spin_reset_rand_N6_z_V10	spin z	60	30	2.61
spin_reset_rand_N6_z_V5	spin z	30	30	2.75
spin_reset_rand_N6_z_edge	spin z	6	30	2.99
spin_reset_rand_N6_z_frozen	spin z	6	30	3.00
spin_reset_rand_N6_z_pm	spin z	6	30	2.98
spin_reset_rand_N6_z_pm_slow	spin z	6	30	2.98
spin_reset_rand_N6_z_scrambled	spin z	6	30	2.92
spin_reset_rand_N6_z_weak0.1	spin online weak	6	1	2.99
spin_reset_rand_N6_z_weak0.3	spin online weak	6	1	2.99
spin_reset_rand_N6_z_weak0.6	spin online weak	6	1	2.98
spin_reset_rand_N6_zz	spin z+zz	21	30	2.72
spin_reset_rand_N6_zz_V10	spin z+zz	210	30	2.19
spin_reset_rand_N6_zz_pm	spin z+zz	21	30	2.73
spin_reset_rand_N6_zz_weak0.1	spin online weak	21	1	2.72
spin_reset_rand_N6_zz_weak0.3	spin online weak	21	1	2.70
spin_reset_rand_N6_zz_weak0.6	spin online weak	21	1	2.68
spin_reset_rand_N8_z	spin z	8	30	2.99

1.06/1.79/1.48/3.19/1.04 (spin z, zz, xyz, Gaussian, CV Kerr) and the accurate-cell medians to 59/129/315/30/2.3, so every conclusion of Sec. III.B holds under the charged convention.

[1] H. Jaeger and H. Haas, Science **304**, 78 (2004).
[2] L. Appeltant, M. C. Soriano, G. Van der Sande, J. Danckaert, S. Massar, J. Dambre, B. Schrauwen, C. R. Mirasso, and I. Fischer, Nature Communications **2**, 468 (2011).
[3] G. Tanaka, T. Yamane, J. B. Héroux, R. Nakane, N. Kanazawa, S. Takeda, H. Numata, D. Nakano, and A. Hirose, Neural Networks **115**, 100 (2019).
[4] K. Fujii and K. Nakajima, Physical Review Applied **8**, 024030 (2017).
[5] P. Mujal, R. Martínez-Peña, J. Nokkala, J. García-Beni, G. L. Giorgi, M. C. Soriano, and R. Zambrini, Advanced Quantum Technologies **4**, 2100027 (2021).
[6] S. Ghosh, A. Opala, M. Matuszewski, T. Paterek, and T. C. H. Liew, npj Quantum Information **5**, 35 (2019).
[7] Y. Hou, J. Hua, Z. Wu, W. Xia, Y. Chen, X. Li, Z. Li, X. Peng, and J. Du, High-accuracy temporal prediction via experimental quantum reservoir computing in correlated spins (2025), arXiv:2508.12383.
[8] A. Senanian, S. Prabhu, V. Kremenetski, S. Roy, Y. Cao, J. Kline, T. Onodera, L. G. Wright, X. Wu, V. Fatemi, and P. L. McMahon, Nature Communications **15**, 7490 (2024).
[9] M. Kornjača, H.-Y. Hu, C. Zhao, J. Wurtz, P. Weinberg, M. Hamdan, A. Zhdanov, S. H. Cantu, H. Zhou, R. Araiza Bravo, K. Bagnall, J. I. Basham, J. Campo, A. Choukri, R. DeAngelo, P. Frederick, D. Haines, J. Hammett, N. Hsu, M.-G. Hu, F. Huber, P. N. Jepsen, N. Jia, T. Karolyshyn, M. Kwon, J. Long, J. Lopatin, A. Lukin, T. Macrì, O. Marković, L. A. Martínez-Martínez, X. Meng, E. Ostroumov, D. Paquette, J. Robinson, P. Sales Rodriguez, A. Singh, N. Sinha, H. Thoreen, N. Wan, D. Waxman-Lenz, T. Wong, K.-H. Wu, P. L. S. Lopes, Y. Boger, N. Gemelke, T. Kitagawa, A. Keesling, X. Gao, A. Bylinskii, S. F. Yelin, F. Liu, and S.-T. Wang, Large-scale quantum reservoir learning with an analog quantum computer (2024), arXiv:2407.02553.
[10] V. Cimini *et al.*, npj Quantum Information (2026).
[11] T. Yasuda, Y. Suzuki, T. Kubota, K. Nakajima, Q. Gao, W. Zhang, S. Shimono, A. Matsuo, J. Ananthanarayanan, H. I. Nurdin, and N. Yamamoto, Quantum reservoir computing with repeated measurements on superconducting devices (2023), arXiv:2310.06706.
[12] P. H. Nguyen and D. Soh, Quantum-Fisher-information-aware optimization yields a quantum reservoir advantage over tuned classical reservoirs (2026), arXiv:2610.14454.
[13] T. Pandey, When classical baselines are tuned as carefully as the quantum model, does quantum reservoir computing still win? (2026), arXiv:2607.09905.
[14] J. Connerty *et al.*, Regularisation closes the gap between quantum and classical reservoirs (2026).
[15] S. Tariq, M. Talha, A. Ali, M. Diyan, and S. Chatzinotas, Quantum reservoir computing: Recent advances and future directions (2026), arXiv:2607.18552.
[16] H. Zadeh-Haghighi and C. Simon, Quantum reservoir computing for attention-like selective reproduction (2026), arXiv:2610.19272.
[17] D. Soh, Computational superiority of non-Markovian Kerr feedback in continuous-variable quantum reservoir computing (2026), arXiv:2606.06689.
[18] J. Nokkala, F. Centrone, and F. Arzani, A hidden bottleneck in classical and quantum linear reservoir computing (2026), arXiv:2605.29071.
[19] W. Xia, S. Cao, X. Qiu, and X. Li, Quantum magic and non-commutativity as computational resources in quantum reservoir computing (2026), arXiv:2607.12035.
[20] J. Freiheit and F. Campaioli, Superextensive learning in quantum reservoirs at the onset of information scrambling (2026), arXiv:2608.25511.
[21] S. Cao, W. Xia, and X. Li, Theory of the edge-of-chaos advantage in quantum reservoir computing (2026), arXiv:2610.05213.
[22] R. Martínez-Peña, G. L. Giorgi, J. Nokkala, M. C. Soriano, and R. Zambrini, Physical Review Letters **127**, 100502 (2021).
[23] D. Soh, A charge selection rule fixes what a squeezed-light reservoir computer can compute and afford (2026), arXiv:2608.19668.
[24] P. Mujal, R. Martínez-Peña, G. L. Giorgi, M. C. Soriano, and R. Zambrini, npj Quantum Information **9**, 16 (2023).
[25] A. Sannia *et al.*, npj Quantum Information (2026).
[26] J. Nokkala, R. Martínez-Peña, G. L. Giorgi, V. Parigi, M. C. Soriano, and R. Zambrini, Communications Physics **4**, 53 (2021).
[27] A. F. Atiya and A. G. Parlos, IEEE Transactions on Neural Networks **11**, 697 (2000).
[28] M. C. Mackey and L. Glass, Science **197**, 287 (1977).
[29] E. N. Lorenz, Journal of the Atmospheric Sciences **20**, 130 (1963).
[30] H. Jaeger, *Short term memory in echo state networks*, Tech. Rep. GMD Report 152 (GMD – German National Research Institute for Computer Science, 2002).
[31] S. Aaronson and A. Ambainis, in *Proceedings of the 47th Annual ACM Symposium on Theory of Computing (STOC)* (2015) pp. 307–316.
[32] U. Girish and R. Servedio, Forrelation is extremally hard (2025), arXiv:2508.02514.
[33] R. Gold, IEEE Transactions on Information Theory **14**, 154 (1968).
[34] O. S. Rothaus, Journal of Combinatorial Theory, Series A **20**, 300 (1976).
[35] S. Bravyi, D. Gosset, D. Grier, and L. Schaeffer, Quantum **5**, 567 (2021), arXiv:2102.06963.
[36] A. Coll-Comas *et al.*, Hybrid reservoir computing with quantum feature maps and classical echo-state post-processing (2026).
[37] H. Wakaura, qrcfit: a resource map of quantum reservoir computing against tuned classical baselines (manuscript, code and results, v0.1.1), Version 0.1.1, [Dataset and Software], Zenodo, `https://doi.org/10.5281/zenodo.23663395`; source: `https://github.com/deeptell-inc/qrcfit` (2026).