

Diagnosing Quantum Reservoirs at scale based on expressivity and coverage

Laia Domingo^{1,2}, Oriol Balló-Gimbernat^{3,1,2}, Fernando Vilariño^{1,2}

¹Centre de Visió per Computador (CVC), Barcelona, Spain

²Universitat Autònoma de Barcelona (UAB), Barcelona, Spain

³Eurecat, Centre Tecnològic de Catalunya, Barcelona, Spain

Quantum reservoirs (QRs) are among the most hardware-friendly approaches to quantum machine learning on near-term devices. Inherited from classical reservoir computing [1–4], their core idea is to use the dynamics of a fixed, typically random quantum system to process input data. Each data point is embedded into the reservoir state, evolved through the quantum dynamics, and represented by a set of measured observables. These observables form a feature vector that is passed to a classical readout, usually a simple linear regression, under the premise that the reservoir dynamics have extracted sufficiently informative features for the target task.

The same mechanism appears in two related settings. In quantum reservoir computing (QRC), the reservoir processes temporal signals and retains a fading memory of past inputs, making it suitable for time-series tasks [5–7]. In the quantum extreme learning machine (QELM) setting, inputs are processed independently, with no memory across samples, which is suitable for static regression and classification [8]. Since the quantum system itself contains no trainable parameters, the QR performance depends strongly on the choice of its fixed random dynamics. At scale, QR diagnostics become design tools, identifying which dynamics are worth implementing and likely to yield usable predictive power. Several diagnostics have therefore been proposed to characterize good reservoirs. Dynamical and spectral criteria relate performance to quantum chaos and to the transition between integrable and chaotic regimes, for instance through level-spacing statistics in transverse-field Ising models [9]. The majorization principle [10, 11] ranks random circuits by the statistical disorder of their output distributions and has been shown to correlate with QR performance [12]. Other approaches, such as the entanglement spectrum [13] and Krylov complexity [14], quantify the dynamical complexity of the QR through the spreading of quantum states.

These indicators, however, do not provide a fully general and scalable diagnostic for QRs. Some are tied to specific physical models and do not naturally apply to arbitrary analog or gate-based reservoirs. Others are more general but require resources that grow prohibitively with system size: they require access to the full state, the full unitary, or the full output distribution. What is missing is a diagnostic that is general across reservoir families, scalable beyond small Hilbert spaces, and robust to finite sampling and hardware noise. In this work, we propose two complementary diagnostics that address this gap. The first is a scalable measure of *expressivity* based on the order-statistics (ORS) score, recently introduced for simulation-free fidelity estimation of individual circuits [15]. We adapt ORS to quantify how close a reservoir ensemble is to Haar-random behavior through its measurement statistics. Instead of reconstructing a full distribution, ORS uses only the few most likely measurement outcomes, giving a cost that is independent of Hilbert-space dimension. Crucially, we show that both diagnostics are strongly predictive of learning performance across QELM and QRC tasks, so they can guide reservoir design rather than only characterize it after training.

Let a reservoir instance produce probabilities $\{p_j\}_{j=1}^D$ over the $D = 2^n$ computational-basis states. We sort them as

$$p_{(1)} \geq p_{(2)} \geq \dots \geq p_{(D)}. \quad (1)$$

For a Haar-random state, the distribution $P_k(x)$ of the k -th largest probability is known analytically [15]. Given an ensemble of M reservoir instances, we evaluate the Haar log-likelihood of the observed top- K ranks,

$$\bar{\Lambda}_K = \frac{1}{M} \sum_{m=1}^M \sum_{k=1}^K \frac{1}{k} \log P_k(p_{(k)|m}), \quad (2)$$

where $p_{(k)|m}$ is the k -th largest output probability of the m -th reservoir instance. The Haar reference value, denoted $\Lambda_{\text{Haar},K}$, is obtained by evaluating the same quantity analytically under the Haar order-statistics distribution. We then define the ORS expressivity diagnostic as the signed gap

$$G_{\text{ORS},K} = \bar{\Lambda}_K - \Lambda_{\text{Haar},K}. \quad (3)$$

Reservoir ensembles whose top-ranked output probabilities follow Haar order statistics have $G_{\text{ORS},K} \simeq 0$, while deviations from zero indicate departures from Haar-random behavior. This provides a scalable diagnostic of reservoir expressivity that is defined for arbitrary reservoir families and depends only on K measured bitstrings per reservoir instance, with $K \ll D$. The ORS diagnostic has a natural extension to depolarizing noise [15]. If f denotes an effective circuit fidelity, the noisy probability and noisy ORS are defined as

$$p_j^{(f)} = f p_j + \frac{1-f}{D}, \quad P_k^{(f)}(x) = \frac{1}{f} P_k\left(\frac{x - (1-f)/D}{f}\right). \quad (4)$$

For $f > 0$, this affine transformation preserves the ordering of the probabilities. The Haar reference shifts by $\Lambda_{\text{Haar},K}^{(f)} = \Lambda_{\text{Haar},K} - \sum_{k=1}^K \frac{1}{k} \log f$, so under exact global depolarizing the corrected gap $G_{\text{ORS},K}^{(f)}$ is invariant in f : gaps at different fidelities are directly comparable, and ORS is usable in noisy simulation and on hardware. We also introduce a multi-basis extension. A truly Haar-random ensemble is basis invariant, so its statistics should follow Haar order statistics not only in the computational basis but after fixed basis changes. For a set $\mathcal{B}$ of random local Pauli measurement bases, we compute the ORS log-likelihood in each basis and average the signed gap of Eq. 3, giving a stricter, more basis-robust expressivity test.

Expressivity alone, however, does not depend on the input distribution, the chosen observables, or the feature matrix seen by the readout. We therefore complement expressivity with a task-dependent diagnostic of *coverage*, defined directly from the feature matrix. Given a training set $\{x_i\}_{i=1}^{N_{\text{train}}}$, let $\mathbf{h}(x_i)$ be the vector of measured reservoir observables for input x_i . We collect these vectors into the feature matrix H and denote its column-centred version by $H_c = H - \bar{H}$. If $\{\sigma_i\}$ are the singular values of H_c , we define the effective rank as the participation ratio

$$R_{\text{eff}}(H_c) = \frac{(\sum_i \sigma_i)^2}{\sum_i \sigma_i^2}. \quad (5)$$

This quantity estimates how many independent feature directions are effectively available to the linear readout. ORS tests whether the reservoir is intrinsically expressive; R_{eff} tests whether this expressivity becomes usable input-dependent information. We evaluate these diagnostics on reservoir families spanning different levels of structural complexity. The first set consists of random circuits generated by gates drawn from

$$G_1 = \{\text{CNOT}, H, X\}, \quad G_2 = \{\text{CNOT}, H, S\}, \quad G_3 = \{\text{CNOT}, H, T\}. \quad (6)$$

The families G_1 and G_2 are non-universal and classically simulable, while the non-Clifford T gate makes G_3 approximately universal as depth increases [16–20]. The second set consists of diagonal IQP-style ensembles [21–23]. The base family D_2 is

$$U = H^{\otimes n} \exp\left(i \sum_q \phi_q Z_q + i \sum_{(u,v) \in E} \phi_{uv} Z_u Z_v\right) H^{\otimes n}, \quad (7)$$

with phases sampled uniformly and E sampled from an Erdős–Rényi graph. Although increasing the mean degree makes D_2 more random-looking in output statistics, all generators commute. To test whether breaking this structure improves coverage, we also consider $D_{2,XZ}$ and $D_{2,XZY}$, obtained by stacking diagonal blocks in different Pauli bases.

We validate ORS against KL expressivity, level-spacing statistics, and Krylov complexity in the small-system regime, where all diagnostics can be computed reliably. All indicators identify the same transition from shallow structured circuits to Haar-like, chaotic behavior as the depth of G_3 increases. We then show that the noise-corrected ORS gap preserves the expressivity ranking under simulated depolarizing noise and remains informative in experiments on IBM quantum hardware. Finally, we relate expressivity and coverage to predictive performance in QELM and QRC benchmarks, including synthetic Fourier regression, LiH excited-energy prediction, NARMA time-series prediction, and EMSIG residential energy forecasting [24]. For the synthetic Fourier regression, given a scalar input $x \in [0, 2\pi]$, the target function and input encoding are

$$y(x) = \sum_{j=0}^{J_F} [a_j \cos(jx) + b_j \sin(jx)], \quad |\psi(x)\rangle = \bigotimes_{q=0}^{k_{\text{exp}}-1} R_z(-3^q x) H |0\rangle, \quad (8)$$

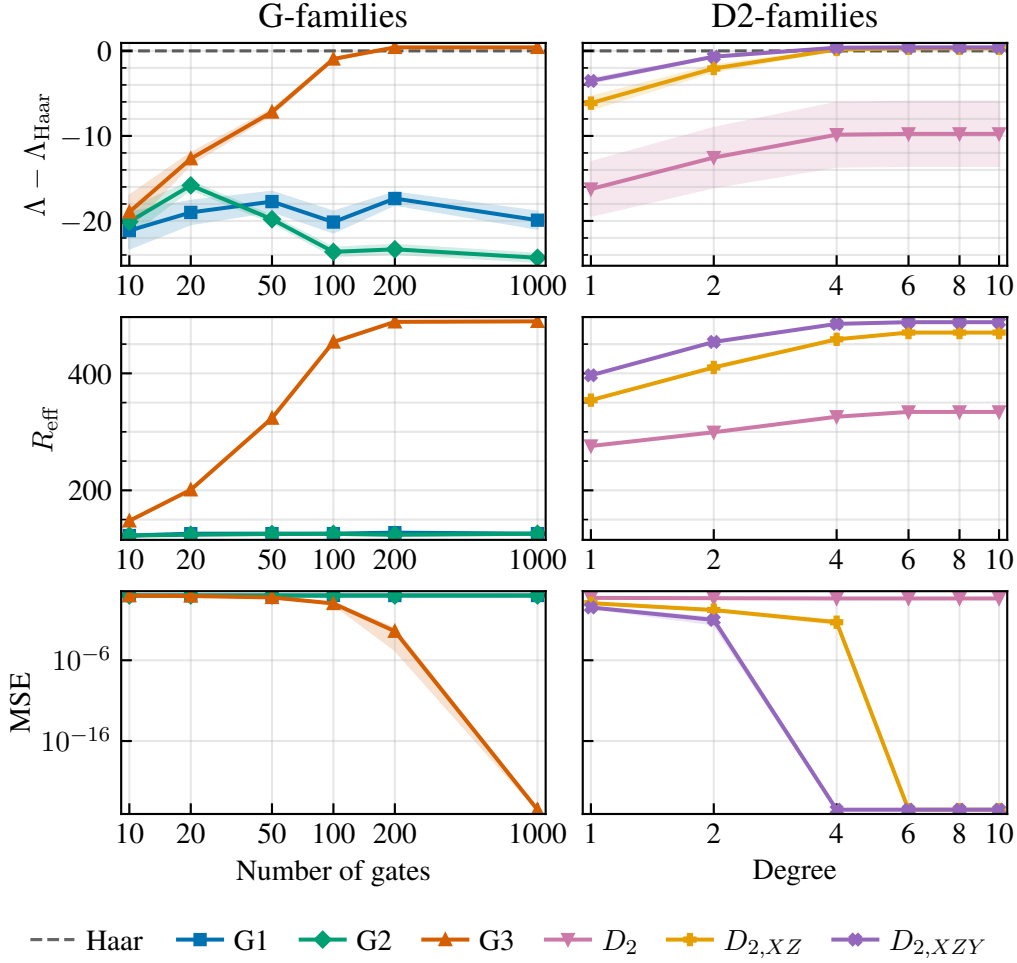


Figure 1: QELM results on the synthetic Fourier regression task for $n = 6$. Top: ORS diagnostic, using the single-basis gap for the G families and the multi-basis diagnostic with $B = 50$ random local Pauli bases for the diagonal families. Middle: effective rank R_{eff} . Bottom: test mean squared error (MSE).

with $J_F = (3^{n_a} - 1)/2$, $n_a = 6$ and a_j, b_j sampled uniformly in $[-1, 1]$. This makes the target Fourier bandwidth compatible with the frequencies accessible to the encoding, so that performance differences can be attributed primarily to the reservoir feature map. We use $n = 6$ qubits and 5000 samples, with 3500 for training and 1500 for testing. The readout uses a fixed pool of $n_{\text{obs}} = 1000$ random Pauli observables shared across all reservoir families.

Across tasks, the expressivity and coverage diagnostics are strongly aligned with performance, as shown in Fig. 1. For the random-gate families, this behavior is clearest in G_3 : as the number of gates increases, G_3 approaches the Haar reference in ORS, its effective rank rises sharply, and the test error decreases. By contrast, G_1 and G_2 remain structurally restricted, do not reach the same Haar-like regime, have lower effective rank, and perform worse. For the IQP-style families, the purely commuting D_2 family is separated from its non-commuting extensions. It remains further from Haar-like statistics across measurement bases, has the lowest effective rank, and gives the weakest predictive performance. Adding non-commuting blocks in $D_{2,XZ}$ and $D_{2,XZY}$ makes the output statistics more basis-robust, increases R_{eff} , and improves test error.

These results support a two-level diagnosis of quantum reservoirs. Multi-basis ORS provides a scalable, task-independent test of intrinsic expressivity, while R_{eff} determines whether that expressivity becomes usable task-dependent information for the readout. Good performance arises when both are favorable. Taken together, these results position ORS and R_{eff} as practical diagnostics for designing QRs rather than merely evaluating them after training. The framework is general across reservoir families, scalable with Hilbert-space dimension, and compatible with noisy hardware. It therefore provides a common language for comparing gate-based, analog, and hardware-native reservoirs, and for selecting reservoirs that are not only expressive but also task-suitable and resource-efficient.

References

- [1] H. Jaeger and H. Haas. Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *Science*, 304:78–80, 2004.
- [2] W. Maass and H. Markram. On the computational power of circuits of spiking neurons. *Journal of Computer and System Sciences*, 69:593–616, 2004.
- [3] G. Tanaka, T. Yamane, J. B. Héroux, R. Nakane, N. Kanazawa, S. Takeda, H. Numata, D. Nakano, and A. Hirose. Recent advances in physical reservoir computing: A review. *Neural Networks*, 115:100–123, 2019.
- [4] K. Nakajima. Physical reservoir computing—an introductory perspective. *Japanese Journal of Applied Physics*, 59:060501, 2020.
- [5] J. Chen, H. I. Nurdin, and N. Yamamoto. Temporal information processing on noisy quantum computers. *Phys. Rev. Applied*, 14, 2020.
- [6] K. Kobayashi, K. Fujii, and N. Yamamoto. Feedback-driven quantum reservoir computing for time-series analysis. *PRX Quantum*, 5, 2024.
- [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. *Phys. Rev. Lett.*, 136:120602, 2026. doi: 10.1103/r8ww-qw7j.
- [8] W. Xiong, G. Facelli, M. Sahebi, O. Agnel, T. Chotibut, S. Thanasilp, and Z. Holmes. On fundamental aspects of quantum extreme learning machines. *Quantum Machine Intelligence*, 7:20, 2025. doi: 10.1007/s42484-025-00258-4.
- [9] R. Martínez-Peña, G. L. Giorgi, J. Nokkala, M. C. Soriano, and R. Zambrini. Dynamical phase transitions in quantum reservoir computing. *Phys. Rev. Lett.*, 127:100502, 2021. doi: 10.1103/PhysRevLett.127.100502.
- [10] J. I. Latorre and M. A. Martín-Delgado. Majorization arrow in quantum-algorithm design. *Phys. Rev. A*, 66:022305, 2002.
- [11] Raúl O. Vallejos, Fernando de Melo, and Gabriel G. Carlo. Principle of majorization: Application to random quantum circuits. *Phys. Rev. A*, 104:012602, Jul 2021. doi: 10.1103/PhysRevA.104.012602. URL <https://link.aps.org/doi/10.1103/PhysRevA.104.012602>.
- [12] L. Domingo, G. Carlo, and F. Borondo. Optimal quantum reservoir computing for the noisy intermediate-scale quantum era. *Phys. Rev. E*, 106:L043301, 2022. doi: 10.1103/PhysRevE.106.L043301.
- [13] D. Shaffer, C. Chamon, A. Hamma, and E. R. Mucciolo. Irreversibility and entanglement spectrum statistics in quantum circuits. *Journal of Statistical Mechanics: Theory and Experiment*, 2014:P12007, 2014.
- [14] L. Domingo, F. Borondo, G. Scialchi, A. J. Roncaglia, G. G. Carlo, and D. A. Wisniacki. Quantum reservoir complexity by the krylov evolution approach. *Phys. Rev. A*, 110:022446, 2024. doi: 10.1103/PhysRevA.110.022446.
- [15] Tobias Micklitz. Simulation-free fidelity estimation via quantum output order statistics. 2025.
- [16] D. Gottesman. The heisenberg representation of quantum computers. 1998.
- [17] S. Clark, R. Jozsa, and N. Linden. Generalized clifford groups and simulation of associated quantum circuits. *Quantum Information & Computation*, 8:106–126, 2008.
- [18] M. Van Den Nest. Classical simulation of quantum computation, the Gottesman-Knill theorem, and slightly beyond. *Quantum Information & Computation*, 10:258–271, 2010.

- [19] R. Jozsa and M. Van Den Nest. Classical simulation complexity of extended clifford circuits. *Quantum Information & Computation*, 14:633–648, 2014.
- [20] D. E. Koh. Further extensions of clifford circuits and their classical simulation complexities. *Quantum Information & Computation*, 17:262–282, 2017.
- [21] M. J. Bremner, R. Jozsa, and D. J. Shepherd. Classical simulation of commuting quantum computations implies collapse of the polynomial hierarchy. *Proc. R. Soc. A*, 467:459, 2011.
- [22] X. Ni and M. Van Den Nest. Commuting quantum circuits: Efficient classical simulations versus hardness results. *Quantum Information & Computation*, 13:54–72, 2013.
- [23] K. Fujii and T. Morimae. Commuting quantum circuits and complexity of ising partition functions. *New Journal of Physics*, 19:033003, 2017.
- [24] K. Brucke, S. Schmitz, D. Köglmayr, S. Baur, C. R  th, E. Ansari, and P. Klement. Benchmarking reservoir computing for residential energy demand forecasting. *Energy and Buildings*, 314:114236, 2024. doi: 10.1016/j.enbuild.2024.114236.