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.

Benchmarking Non-Markovianity Measures in Digital Quantum Simulation of Open Quantum Systems

S. M. Yousuf Iqbal Tomal

yousuf.iqbal.tomal@g.bracu.ac.bd

BRAC University https://orcid.org/0009-0000-3391-3824

Research Article

Keywords: non-Markovianity, open quantum systems, digital quantum simulation, non-Markovianity measures, pseudomode embedding, NISQ benchmarking

Posted Date: July 21st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10406502/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.

Benchmarking Non-Markovianity Measures in Digital Quantum Simulation of Open Quantum Systems

S.M. Yousuf Iqbal Tomal*
Department of Computer Science and Engineering, BRAC University, Dhaka, Bangladesh

July 18, 2026

Abstract

Digital quantum simulation of open quantum systems has produced a range of methods for reproducing non-Markovian dynamics, including pseudomode embeddings, the hierarchical equations of motion, and collision models. These methods are validated against populations, coherences, or spectra. They are not validated against the quantities that define memory itself, and the behavior of those quantities under the pressures of a real simulation is largely unexamined. We study three such pressures. First, a single-qubit mode encoding misrepresents the Breuer-Laine-Piilo (BLP) measure with an error whose sign flips between regimes: it fabricates memory at strong coupling, by about a factor of 2.6 over the converged value, and it erases memory at moderate temperature. The Rivas-Huelga-Plenio (RHP) measure corroborates the temperature-driven destruction, while at strong coupling it is numerically unstable and does not track the BLP result. We derive a cross-validated rule for the required Fock dimension that collapses both regimes onto the occupied number extent, $d_{\mathrm{req}} \approx \lceil \langle n \rangle + c\,\sigma_n \rceil + 1$, and we show that the constant must be calibrated on super-Poissonian statistics. Second, using the exact noisy measure as computable ground truth, we show that the raw finite-shot estimate of the BLP measure can nearly double the true value under realistic noise. We provide a debiased estimator that is validated near-unbiased in that regime, together with a Markovian null-floor control that ports to hardware. Third, temporally correlated device noise is conflated with physical memory by the measure. A resource-fair comparison finds that a pseudomode and an ancilla-train collision model reach parity at equal qubit cost for a single-Lorentzian bath. Every result is simulator-based, validated against an exact reference, and reproducible from recorded data. We make no claim of quantum advantage.

Keywords: non-Markovianity, open quantum systems, digital quantum simulation, non-Markovianity measures, pseudomode embedding, NISQ benchmarking

1 Introduction

An open quantum system exchanges energy and information with an environment [1, 2]. When the environment retains a memory of past interaction and returns information to the system, the reduced dynamics is non-Markovian [3–5]. Memory of this kind matters for solid-state qubits, for structured photonic environments, and for energy transport in molecular aggregates [6, 7]. Quantifying it precisely is therefore a practical concern, not only a foundational one.

Two measures are standard. The Breuer-Laine-Piilo (BLP) measure detects the backflow of information through revivals of the trace distance between two evolving states [8, 9]. The Rivas-Huelga-Plenio (RHP) measure detects departures from completely positive divisibility of the dynamical map [10]. Other measures quantify memory through correlations, the volume

*Corresponding author: yousuf.iqbal.tomal@g.bracu.ac.bd; ORCiD: 0009-0000-3391-3824.

of accessible states, or channel capacities [11–13], and the relationships among them form a hierarchy [14–16]. All of them were designed for use with exact dynamics or clean experimental data [17].

Simulating non-Markovian dynamics on a quantum computer is difficult for a structural reason. A quantum circuit implements a unitary map, while open system dynamics is non-unitary [18, 19]. Every method for the problem is a different way of embedding the non-unitary evolution into a larger unitary one [20]. The pseudomode approach replaces a structured bath with a small number of damped harmonic oscillators [21–24]. The hierarchical equations of motion propagate a set of auxiliary density operators [25–27], and a recent quantum algorithm dilates that propagator onto a circuit [28]. General algorithms for Lindblad and dissipative evolution on a quantum computer carry rigorous complexity guarantees [29, 30]. Collision models build the environment from a stream of ancillas [31, 32], and ancilla-train and chain constructions extend this to structured baths on hardware [33–36]. Several works implement open-system and non-Markovian dynamics on near-term devices directly [37, 38].

These methods share a common validation practice. They compare simulated populations, coherences, or optical spectra against a reference. They do not compare a rigorous non-Markovianity measure against a reference, and they do not examine how such a measure behaves once the simulation introduces truncation, finite sampling, and device noise. This gap matters, because the BLP measure is a rectifying functional: it sums only the positive parts of a signal, so it converts statistical fluctuation into apparent memory. A measure used as evidence of memory should be shown to report memory faithfully under exactly the conditions in which it will be used, and those conditions on present hardware include shallow-to-moderate depth, finite shots, and correlated noise [39, 40].

We address this gap with a validated, simulator-based framework, summarized in Figure 1. Our contributions are the following.

1. **A measure-first validation chain.** We compute the BLP and RHP measures from a pseudomode circuit and check them against a closed-form solution and a master equation solve, with agreement at the level of 10^{-4} or better (Section 2).

2. **Truncation misrepresents memory, with a regime-dependent sign.** A single-qubit mode encoding inflates the BLP measure at strong coupling and destroys it at moderate temperature; the RHP measure corroborates the finite-temperature case, while it is numerically unstable at strong coupling. We quantify both effects and trace them to the number statistics of the occupied mode (Section 3).

3. **A cross-validated resource rule.** The required Fock dimension follows the occupied number extent. The rule is safe only when its constant is calibrated on the broader, super-Poissonian statistics of a thermal mode (Section 4).

4. **Ground-truth-validated shot statistics.** Because a density-matrix simulator gives the exact noisy measure, we measure the bias of shot-based estimators directly. We provide a debiased estimator and a hardware-portable null-floor control (Section 5).

5. **Device-noise conflation.** Temporally correlated noise preserves the measured memory more than memoryless noise of matched strength, so the measure cannot separate physical memory from noise correlation (Section 6).

6. **A resource-fair embedding comparison.** A pseudomode and an ancilla-train collision model reach parity at equal qubit cost for a single-Lorentzian bath (Section 7).

The framework runs on classical simulators. It contains no quantum-advantage claim. Its value is diagnostic: it states what breaks when memory is simulated, by how much, and which controls restore an honest accounting.

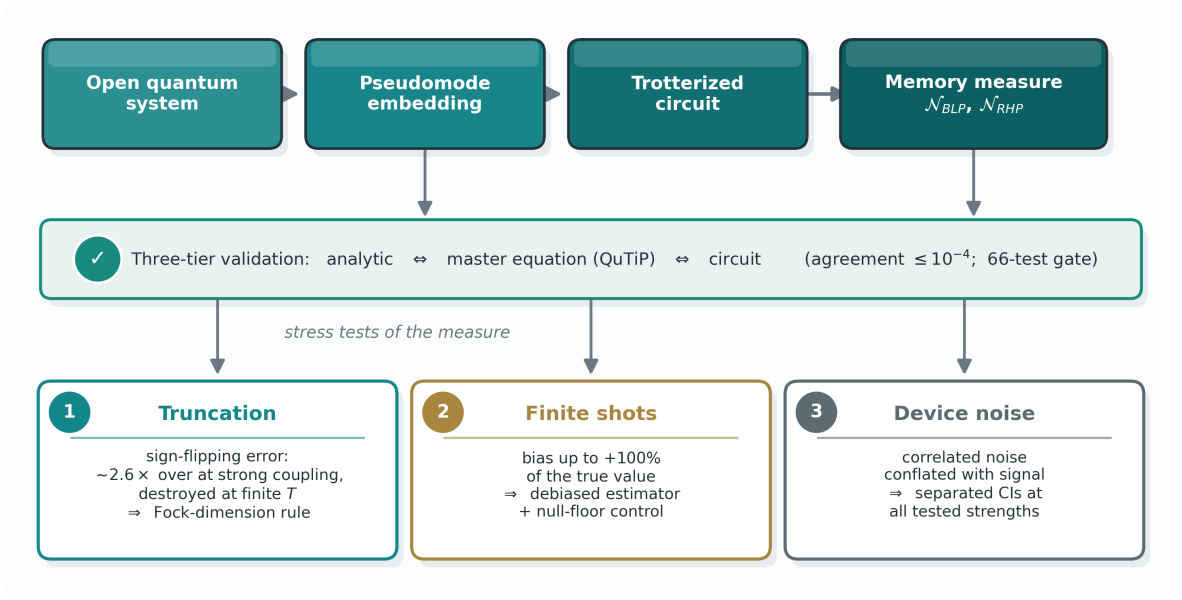


Figure 1: Overview of the framework. A qubit coupled to a Lorentzian bath is mapped to a pseudomode, Trotterized into a circuit, and read out through a non-Markovianity measure. A three-tier chain validates the pipeline against a closed-form solution and a master equation solve. The measure is then stress tested along three axes that any real simulation must face: state-space truncation, finite sampling, and device noise. Each axis produces a specific, quantified failure and a control that addresses it.

2 Framework and validation

2.1 Models

The primary test system, which we call Stage A, is a qubit coupled to a bath with a Lorentzian spectral density. This is the damped Jaynes-Cummings model, and it has a closed-form solution [1]. The reduced system evolves under a time-local master equation whose general form is

$$\dot{\rho} = -\frac{i}{\hbar}[H_S, \rho] + \sum_k \gamma_k \left(L_k \rho L_k^\dagger - \tfrac{1}{2}\{L_k^\dagger L_k, \rho\} \right), \tag{1}$$

with time-dependent rates in the non-Markovian regime [18, 19]. In the single-excitation sector the excited-state amplitude is

$$G(t) = e^{-\lambda t/2} \left[\cosh\frac{dt}{2} + \frac{\lambda}{d} \sinh\frac{dt}{2} \right], \qquad d = \sqrt{\lambda^2 - 2\gamma_0 \lambda}, \tag{2}$$

where λ is the Lorentzian half-width and γ_0 sets the coupling. The dynamics is non-Markovian for

$$\gamma_0 > \lambda/2, \tag{3}$$

where d is imaginary and $G(t)$ oscillates through zero. We use $\gamma_0 = 5$, $\lambda = 1$ for the non-Markovian case and $\gamma_0 = 0.2$ for a Markovian control.

A single Lorentzian maps exactly onto one damped pseudomode [21]. The enlarged system obeys a Markovian instance of Eq. (1) with a coherent exchange between the system qubit and one bosonic mode,

$$H = \Omega\,(\sigma_+ a + \sigma_- a^\dagger), \qquad \Omega = \sqrt{\tfrac{1}{2}\gamma_0 \lambda}, \qquad \kappa = 2\lambda, \tag{4}$$

and a single collapse operator at zero temperature,

$$L_1 = \sqrt{\kappa}\, a. \tag{5}$$

We take $|1\rangle$ to be the excited state throughout, with $\sigma_\pm$ the corresponding raising and lowering operators. Because H conserves excitation number and every initial state we use carries at most one excitation, a two-level truncation of the mode is exact at zero temperature. This is what makes Stage A a controlled testbed: the truncation error there is known to vanish, so any nonzero truncation error elsewhere is attributable to the physics, not to a coding artifact.

At finite temperature the mode relaxes toward a thermal occupation $\bar{n}$ through two channels,

$$L_1 = \sqrt{\kappa(\bar{n}+1)}\, a, \qquad L_2 = \sqrt{\kappa\,\bar{n}}\, a^\dagger. \tag{6}$$

The two-level truncation is then no longer exact, and the required dimension becomes a subject of study rather than an assumption.

We use two further systems. Stage B is a spin-boson model [6] with two damped pseudomodes, which we use to reach strong coupling, with a system Hamiltonian $H_S = \frac{\Delta}{2}\sigma_x$ and a coupling $g\,\sigma_z(a_k + a_k^\dagger)$ to each mode. The excitonic dimer of Section 8 is a two-site energy-transfer unit with one dephasing pseudomode per site.

2.2 Embeddings

The pseudomode embedding Trotterizes Eq. (1) [41, 42]. Figure 2 shows one step. The coherent part is a system-mode exchange, realized by the commuting rotations $R_{XX}(\Omega\,\delta t)\,R_{YY}(\Omega\,\delta t)$. The dissipative part is an amplitude-damping channel, realized by a Stinespring dilation [43, 44]: a controlled rotation onto a fresh ancilla followed by a reset, with per-step damping probability

$$p = 1 - e^{-\kappa\,\delta t}. \tag{7}$$

The construction is deterministic and works directly on a density-matrix simulator. It is non-variational, so it avoids the trainability obstacles that affect variational open-system methods [45]. At finite temperature we split the two channels of Eq. (6) into a generalized amplitude-damping block with emission and absorption probabilities

$$p_\downarrow = 1 - e^{-\kappa(\bar{n}+1)\,\delta t}, \qquad p_\uparrow = 1 - e^{-\kappa\bar{n}\,\delta t}. \tag{8}$$

This block reaches the exact discrete fixed point and converges to the continuous thermal Lindblad equation at first order in δt. The emission-then-absorption ordering within a step is a design choice; other orderings share the same fixed point but carry a different $O(\delta t^2)$ prefactor, so Eq. (8) is one valid discretization rather than the unique one.

The second embedding is an ancilla-train collision model, described in Section 7. It discretizes the bath into a chain of ancillas with a lossy boundary, in the spirit of a chain mapping [35, 36].

2.3 Measures

The trace distance between two states is

$$D(\rho_1, \rho_2) = \tfrac{1}{2}\,\|\rho_1 - \rho_2\|_1 = \tfrac{1}{2}\,\mathrm{Tr}\sqrt{(\rho_1-\rho_2)^\dagger(\rho_1-\rho_2)}. \tag{9}$$

It is a contraction under any completely positive trace-preserving map, so a Markovian process never increases it. The BLP measure is the total increase of this distance, maximized over initial state pairs,

$$\mathcal{N}_{\mathrm{BLP}} = \max_{\rho_1(0),\rho_2(0)} \int_{\sigma>0} \sigma(t)\,dt, \qquad \sigma(t) = \frac{d}{dt}\,D(\rho_1(t),\rho_2(t))\,. \tag{10}$$

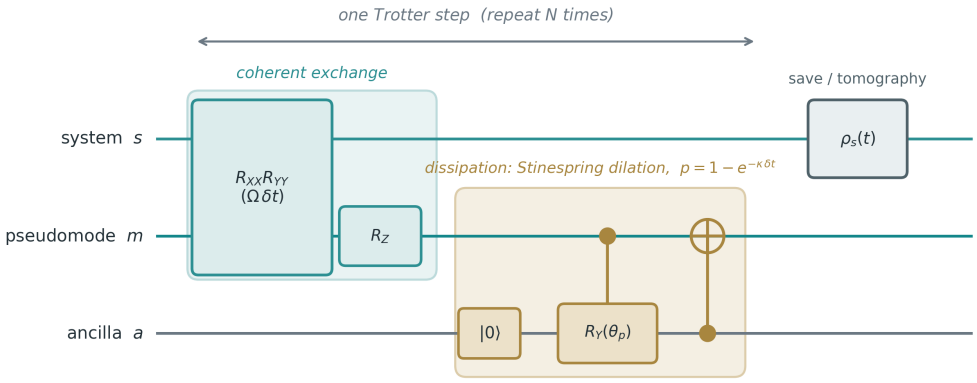


Figure 2: One Trotter step of the pseudomode embedding. The system qubit s and the pseudomode qubit m undergo a coherent exchange $R_{XX}R_{YY}(\Omega\,\delta t)$ and a mode phase R_Z. The dissipation is a Stinespring dilation of the amplitude-damping channel: a fresh ancilla a is reset, a controlled rotation $R_Y(\theta_p)$ with $\theta_p = 2\arcsin\sqrt{p}$ and p from Eq. (7) entangles it with the mode, and a CNOT completes the map. The step repeats N times; the system reduced state $\rho_s(t)$ is read out by a saved density matrix (exact mode) or by Pauli tomography (shot mode).

On a discrete time grid $\{t_j\}$ we evaluate

$$\mathcal{N}_{\mathrm{BLP}} = \max_{\mathrm{pair}} \sum_{j:\,\Delta D_j>0} \Delta D_j, \qquad \Delta D_j = D(t_{j+1}) - D(t_j), \tag{11}$$

which sums only the positive increments. Optimal BLP pairs are orthogonal boundary states [46], which for a qubit are antipodal pure states on the Bloch sphere. We therefore optimize over the sphere, seeded with the three canonical axes so the result is never below the canonical value, with a fixed search budget stated in Appendix B. For Stage A the optimal pair is known analytically and the canonical value is exact. For other models we report the optimized value, and where a coarser canonical estimate is used we label it as a lower bound.

The RHP measure is computed from the reconstructed dynamical map $\Lambda(t)$. We form the intermediate map $\Lambda(t+\delta t)\Lambda(t)^{-1}$, take its Choi matrix C, and integrate the trace-norm excess above unity,

$$\mathcal{N}_{\mathrm{RHP}} = \int g(t)\,dt, \qquad g(t) = \|C(t)\|_1 - 1. \tag{12}$$

The intermediate map is singular where $|G(t)| = 0$, which is exactly the deep non-Markovian regime. We regularize by forming $\Lambda(t)^{-1}$ as a Moore-Penrose pseudoinverse with a singular-value cutoff, discarding singular values below a relative threshold $r_{\mathrm{cond}} = 10^{-6}$ of the largest, rather than a Tikhonov-damped inverse. A step is flagged ill-conditioned when the ratio of the smallest to the largest singular value of $\Lambda(t)$ falls below that threshold, and we report the fraction of flagged steps alongside every RHP value so the reader can judge how much of the integral rests on regularized inversions.

2.4 The validation chain

Every claim rests on a three-tier chain of agreement. For Stage A we compare the closed-form solution of Eq. (2), a QuTiP master equation solve on the enlarged space [47, 48], and the Qiskit circuit [49]. The master equation reproduces the analytic dynamics to a maximum deviation of 2.7×10^{-7}. The circuit reproduces it to 6.25×10^{-4} at Trotter step $\delta t = 0.05$, with a corresponding BLP error of 7.5×10^{-4}. The Trotter error scales as $\delta t^{1.99}$, and the maximum trajectory deviation falls by a clean factor near four for each halving of the step (9.8×10^{-3}, 2.5×10^{-3}, 6.3×10^{-4},

1.6×10^{-4} at $\delta t = 0.2, 0.1, 0.05, 0.025$), so the second-order behavior is a genuine scaling and not a fit artifact. A first-order Lie-Trotter product formula guarantees only $O(\delta t)$ error in general [41, 42], so the observed order deserves a mechanism. Two features of this model supply it. First, the two coherent generators R_{XX} and R_{YY} commute, so the coherent propagation carries no splitting error and is applied exactly; the usual dominant $O(\delta t)$ term, which comes from splitting non-commuting Hamiltonian pieces, is therefore absent. Second, the only remaining error is the interleaving of the exact coherent step with the dissipative block, and that block acts on the pseudomode while the measured observable is the system reduced state, so the residual splitting error is generated on the mode and reaches the system only through a further coherent exchange, one additional order in δt. We report $\delta t^{1.99}$ as the empirically observed order for this model and observable; the guaranteed bound remains first order.

A test suite of 68 checks enforces these tolerances, including the finite-temperature fixed point, the numerical equivalence of the collision-model engines to 10^{-13}, and the convergence of every circuit to its reference. We treat the suite as a gate: if it fails, downstream results are void until the failure is understood.

Figure 3 shows the chain for the optimal Stage A pair. Two general properties follow immediately and anchor the rest of the paper. The canonical-pair search reproduces the analytic BLP optimum exactly (0.5256 against 0.5256), and the Markovian control returns zero to machine precision. Circuit resources are linear: the transpiled depth grows with slope 1.00 in the number of Trotter steps, and the two-qubit gate count grows as 200 per pseudomode. The Trotter-error, Fock-truncation, and resource scalings are collected in the supplementary Figure 9 (Appendix D).

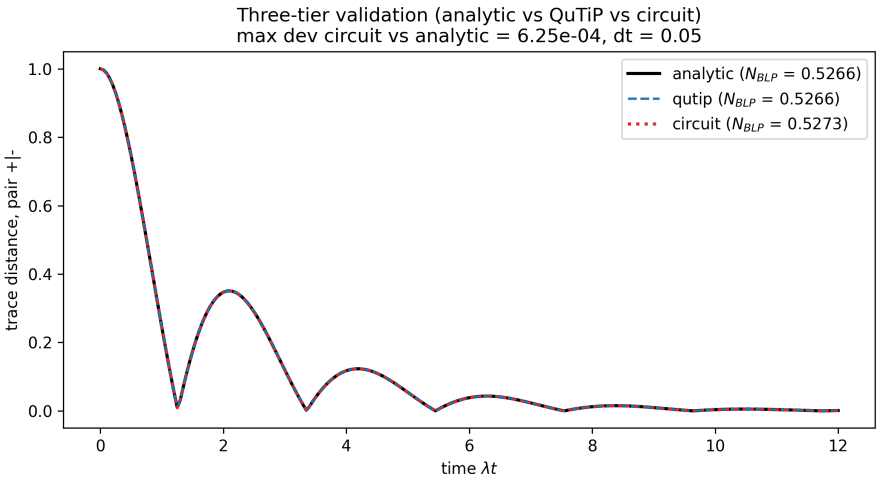


Figure 3: Three-tier validation for Stage A. The trace distance of the optimal pair is shown for the analytic solution, the master equation solve, and the circuit. The circuit reproduces the analytic curve to a maximum deviation of 6.25×10^{-4} at $\delta t = 0.05$, and the three BLP values agree to 7.5×10^{-4}. The Markovian control (not shown) returns zero.

3 Truncation misrepresents memory

The pseudomode is a bosonic mode. A circuit encodes it with a finite number of qubits, which truncates the Fock space. The cheapest encoding uses a single qubit, a two-level truncation. We now show that this encoding misrepresents the BLP measure, and that the sign of the error depends on the regime. Throughout this section the absolute BLP values outside Stage A are canonical-pair lower bounds, and the truncation evidence rests on the pair-independent trajectory deviation, which we report alongside. Tables 1 and 2 give the two regimes numerically, and Figure 4 summarizes both in a single plane.

3.1 Finite temperature: the measure is destroyed

At finite temperature the mode acquires a thermal population, so it explores Fock states above the first excitation. A two-level truncation cannot hold that population. Table 1 shows the consequence. As the thermal occupation $\bar{n}$ increases, the converged reference retains a substantial memory signal, while the two-level circuit collapses toward zero. At $\bar{n} = 1.0$ the circuit reports $\mathcal{N}_{\mathrm{BLP}} = 0.0001$ against a converged value of 0.186. At $\bar{n} = 1.5$ it reports zero against 0.210. The trajectory truncation error grows monotonically from 0.056 to 0.390.

Table 1: Finite-temperature Stage A. The two-level circuit reproduces its matched-truncation reference but diverges from the converged ($n_{\mathrm{Fock}} = 6$) value as temperature rises. Grid $\delta t = 0.05$, $t_{\max} = 10$. Absolute values are canonical-pair lower bounds; the trajectory error is pair-independent.

$\bar{n}$	circuit	ref. (n=2)	converged (n=6)	traj. error
0.00	0.5247	0.5240	0.5240	0.000
0.10	0.3672	0.3679	0.3590	0.056
0.25	0.2152	0.2158	0.2336	0.126
0.50	0.0807	0.0810	0.1439	0.215
0.75	0.0173	0.0167	0.1663	0.280
1.00	0.0001	0.0001	0.1861	0.328
1.50	0.0000	0.0000	0.2102	0.390

Two points make this a diagnostic result rather than a coding artifact. The circuit tracks its matched-truncation reference to within 10^{-2} at every temperature, so the circuit is correct; the fault lies in the truncation it is asked to represent. This also isolates the cause from the generalized amplitude-damping block of Eq. (8). The matched-truncation reference is the exact continuous thermal Lindblad solution at the same two-level cutoff, computed in QuTiP with no discrete damping channel, and the circuit that carries the generalized amplitude-damping block reproduces it to 10^{-2}. The block therefore reaches the correct thermal fixed point at its stated first order in δt, and the collapse is a property of the two-level truncation of the Fock space rather than an artifact of the discrete channel. And at very low temperature the sign is already interesting: at $\bar{n} = 0.1$ the two-level value slightly exceeds the converged value (0.368 against 0.359), a small overestimate that precedes the collapse.

3.2 Strong coupling: the measure is fabricated

The opposite failure appears at strong coupling in Stage B. Here the σ_z coupling does not conserve excitation number, so the mode is driven to higher occupation as the coupling grows. A two-level truncation reflects population that should escape to higher levels, and the reflected population returns to the system as spurious backflow. Table 2 shows the effect. At coupling $g = 0.7$ the two-level circuit reports $\mathcal{N}_{\mathrm{BLP}} = 0.309$ against a converged value of 0.119, an overestimate of about a factor of 2.6. The converged reference is stable: the value at $n_{\mathrm{Fock}} = 6$ and $n_{\mathrm{Fock}} = 7$ differ by only 0.002.

The Stage B bath is defined by its two-pseudomode decomposition rather than fit to an independent physical spectral density, so the truncation-error magnitudes in Table 2 are properties of this pseudomode family; an Ohmic or super-Ohmic continuum mapping would in general give different magnitudes, though the qualitative sign of the error is set by the mode occupation and is expected to persist.

The two regimes together establish the point. The truncation error is not a uniform loss of accuracy. It fabricates memory where the mode is coupling driven, and it destroys memory where the mode is temperature driven. A practitioner who reads a memory value off a two-level

Table 2: Strong-coupling Stage B. The two-level truncation overestimates the measure. Converged reference at $n_{\mathrm{Fock}} = 6$; a 6-versus-7 stability check differs by 0.002 at $g = 0.7$.

g	circuit	ref. (n=2)	converged (n=6)	traj. error
0.4	0.2186	0.2119	0.2393	0.124
0.5	0.2395	0.2333	0.1729	0.169
0.7	0.3095	0.3093	0.1185	0.198

pseudomode simulation can be wrong in either direction, and the direction depends on the physics being studied.

3.3 The sign-flip as a phase diagram

The two regimes above live in different models, one at zero temperature and one at fixed coupling. To exhibit the sign-flip in a single system, we use a finite-temperature spin-boson, in which each damped pseudomode relaxes toward a thermal occupation $\bar{n}$ through emission $\sqrt{\kappa(\bar{n}+1)}\,a$ and absorption $\sqrt{\kappa\bar{n}}\,a^\dagger$. Sweeping the coupling g and the temperature $\bar{n}$ traces the signed truncation error $\mathcal{N}_{\mathrm{BLP}}(n_{\mathrm{Fock}}{=}2) - \mathcal{N}_{\mathrm{BLP}}(\mathrm{converged})$ of the $(+,-)$ pair over a plane, shown in Figure 4. The error is positive at strong coupling and low temperature, where the two-level mode reflects coupling-driven population and fabricates memory, and it is negative at high temperature, where the mode cannot hold the thermal occupation and destroys memory. A single curve, the zero level, separates the two regions. The companion panel shows the converged memory landscape and confirms that the destruction region is one where the reference carries real memory that the truncation removes. The map uses the canonical $(+,-)$ pair on a $\delta t = 0.1$ grid, so it is internally consistent rather than numerically identical to Tables 1 and 2; as in Section 3.2, the magnitudes are specific to this pseudomode-defined bath while the sign structure is set by the mode occupation.

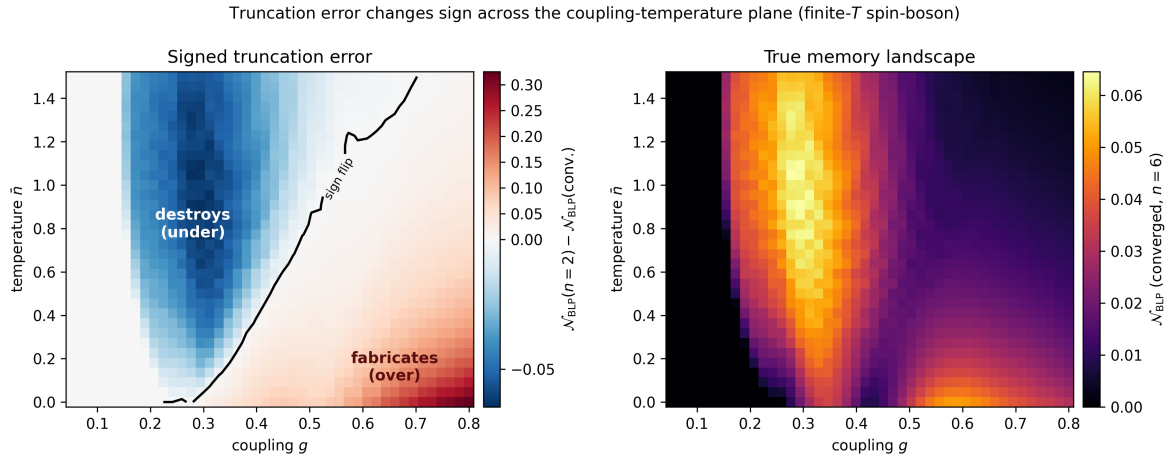


Figure 4: Truncation error across the coupling-temperature plane of a finite-temperature spin-boson. Left: the signed BLP truncation error, $\mathcal{N}_{\mathrm{BLP}}(n{=}2) - \mathcal{N}_{\mathrm{BLP}}(\mathrm{converged})$, with the sign-flip (zero) contour in black. The truncation fabricates memory at strong coupling and low temperature (red) and destroys it at high temperature (blue); the two lobes have genuinely different magnitudes because the converged memory is itself small in the destruction region, so each side of the diverging scale is stretched to fill its half. Right: the converged memory landscape, showing that the destruction region carries real memory that the two-level truncation removes. Values use the $(+,-)$ pair on a $\delta t = 0.1$ grid.

3.4 Corroboration by the RHP measure

The truncation finding above is stated for the BLP measure. To test whether it is specific to that measure, we reconstruct the dynamical map from the four informationally complete inputs and evaluate the RHP measure of Eq. (12) for the two-level circuit and the converged reference. At finite temperature the picture agrees with BLP: the circuit RHP lies below the converged RHP at every temperature (for example 43 against 128 at $\bar{n} = 1$), so both measures report that truncation reduces the memory. The RHP estimator is numerically fragile in this regime, with up to 53% of the intermediate maps ill-conditioned at the $|G(t)| = 0$ crossings, so we read it qualitatively rather than as a precise value. At strong coupling the estimator is numerically unstable: the converged RHP swings non-monotonically with coupling (23, 74, 17, 143 at $g = 0.3, 0.4, 0.5, 0.7$) even though no intermediate map is flagged ill-conditioned there, and at $g = 0.7$ the circuit RHP lies below the converged value while BLP lies above, so the two measures disagree. We attribute this to the sensitivity of the map-reconstruction integral in the strongly dissipative regime, claim RHP corroboration only in the finite-temperature regime, and report the strong-coupling RHP as a numerically unstable null result. The full values are tabulated in Appendix C.

4 A resource rule for the Fock dimension

The natural response is to add Fock levels until the measure converges. How many are needed? We define the required dimension $d_{\mathrm{req}}(\epsilon)$ as the smallest Fock dimension whose reduced dynamics matches a high-dimensional reference to a maximum deviation below ϵ. We compute it across two drivers of occupation: coupling in Stage B at zero temperature, and temperature in Stage A.

A rule based on the mean occupation fails. The two families do not fall on a common curve, and the correlation between d_{req} and the peak mean occupation $\langle n \rangle$ is only 0.78. A thermal mode of a given mean occupation needs more levels than a coupling-driven mode of the same mean. The reason is number statistics. A thermal state is super-Poissonian, with variance

$$\sigma_n^2 = \bar{n}(\bar{n} + 1), \tag{13}$$

while the coupling-driven occupation is closer to Poissonian, with $\sigma_n^2 \approx \langle n \rangle$. Truncation must hold not the mean but the occupied extent, which includes the width.

A variance-aware predictor collapses both families. Using the peak occupation and the number standard deviation at that peak, the rule

$$d_{\mathrm{req}} \approx \lceil \langle n \rangle + c\, \sigma_n \rceil + 1 \tag{14}$$

correlates with the measured d_{req} at 0.938 and provides a safe upper bound with $c = 3$, as shown in Figure 5. The maximum over-provisioning across the tested points is three levels, which occurs at strong coupling.

The calibration of c carries a caveat that a naive fit would hide. We cross-validate over the fourteen grid points, seven temperature and seven coupling, by calibrating c on one family and testing safety on the other. A constant calibrated on the temperature family, $c = 2.75$, remains a safe bound on the coupling family. A constant calibrated on the coupling family, $c = 2.0$, is not safe on the temperature family: it under-provisions three of the seven thermal points, by up to two Fock levels. The rule must therefore be calibrated on the broader, super-Poissonian statistics, or it will under-provision for thermal baths. Appendix B, Table 9 gives the explicit per-family over- and under-provisioning at $c \in \{2.0, 2.75, 3.0\}$. We report the rule across three accuracy targets, $\epsilon \in \{3 \times 10^{-2}, 1 \times 10^{-2}, 3 \times 10^{-3}\}$; at the strictest target one high-temperature point reaches the reference dimension itself, which we flag as right-censored rather than silently clip. The parameter and ϵ domain over which the rule is validated, outside which it is an extrapolation rather than a bound, is tabulated in Appendix B.

We do not present Eq. (14) as a theorem. It is an engineering rule, validated over a stated domain of Lorentzian-type baths, with the number statistics of Eq. (13) giving the physical reason it takes the form it does. Its practical content is direct: a mode of mean occupation near one and thermal width near 1.4 needs about six Fock levels, not two, and a two-level encoding of it will report the wrong memory.

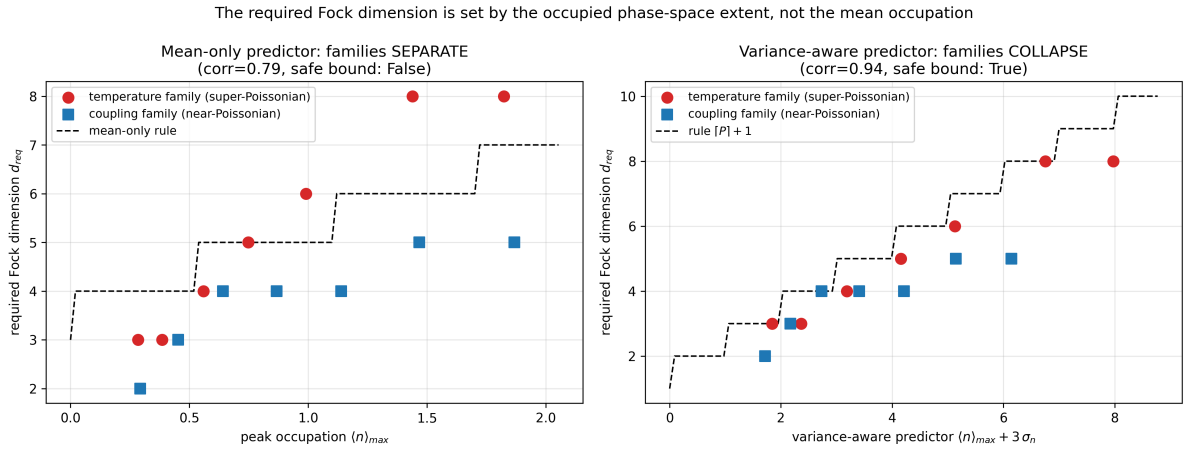


Figure 5: The required Fock dimension. Left: a mean-occupation predictor does not collapse the coupling and temperature families, and a mean-only bound is unsafe. Right: the variance-aware predictor $\langle n\rangle + c\,\sigma_n$ collapses both families onto a single curve, correlation 0.938, and gives a safe bound with $c = 3$. Cross-validation shows the constant must be calibrated on the super-Poissonian thermal family.

5 Finite-shot estimation of the measure

On hardware, and in any honest simulation of a shot-based experiment, the BLP measure is estimated from a finite number of measurements. Equation (11) sums only the positive increments of the trace distance, so it rectifies statistical error into apparent backflow. The size and even the sign of the resulting bias are not obvious. A density-matrix simulator settles the question, because it provides the exact noisy measure as computable ground truth at every noise level. This is the central advantage of a simulator-based study of the statistics, and it turns the following from assertion into measurement.

We scope this finite-shot analysis to the BLP measure deliberately. The BLP measure needs only single-qubit tomography of two propagated states, so its shot cost is modest and its bias is the clean rectification effect studied here. The RHP measure instead requires the full dynamical map at every step, which demands informationally complete process tomography and multiplies the shot budget, and any residual sampling error then passes through the ill-conditioned inversion of Section 3.4, where it is amplified rather than rectified. A shot-resolved RHP estimator is therefore a separate problem; we use RHP only as the exact-state cross-check of Section 3.4 and leave its finite-shot behavior to future work.

We reconstruct each single-qubit state by Pauli tomography. The Bloch vector is

$$\hat{r}_\alpha = \frac{n_0^{(\alpha)} - n_1^{(\alpha)}}{n_0^{(\alpha)} + n_1^{(\alpha)}}, \qquad \alpha \in \{x, y, z\}, \qquad \hat{\rho} = \tfrac{1}{2}\big(\mathbb{1} + \hat{r}_x\sigma_x + \hat{r}_y\sigma_y + \hat{r}_z\sigma_z\big), \tag{15}$$

projected onto the Bloch ball when finite sampling places it outside. We use 4096 shots and 50 random seeds.

5.1 Two biases with opposite signs

Table 3 reports the bias of three estimators against the exact noisy value. Two mechanisms compete. Near the ideal limit the finite-shot Bloch estimate, projected onto the physical ball, smooths and shrinks genuine revivals, so the raw estimate is biased low. Under strong noise the rectification of shot noise dominates, so the raw estimate is biased high. At noise scale two the raw estimate adds 0.064 to a true value of 0.077, an overestimate of about 84%. Reporting a raw shot-based BLP value in that regime is close to meaningless.

Table 3: Estimator bias against the exact noisy $\mathcal{N}_{\mathrm{BLP}}$, averaged over 50 random seeds, 4096 shots, grid $\delta t = 0.15$. Positive bias means overestimate; the standard error of each mean bias is below 0.004. The debiased estimator uses the significance rule of Eq. (16) with $z = 1$.

noise scale	true $\mathcal{N}_{\mathrm{BLP}}$	raw bias	debiased bias	readout-mit. bias
0.0	0.4640	−0.0265	−0.0793	—
0.5	0.2657	+0.0396	−0.0148	+0.0433
1.0	0.1518	+0.0527	−0.0115	+0.0586
2.0	0.0765	+0.0643	−0.0033	+0.0727

We control the rectification bias with a significance threshold. From a bootstrap over the shot counts we obtain the standard error s_j of each trace-distance estimate, and we accept an increment only when it is statistically significant,

$$\mathcal{N}_{\mathrm{BLP}}^{\mathrm{deb}} = \sum_{j:\ \Delta D_j > z\sqrt{s_j^2 + s_{j+1}^2}} \Delta D_j. \tag{16}$$

This estimator is validated near-unbiased in the noise-dominated regime, with a bias of −0.003 at scale two. It over-corrects near the ideal limit, where it discards small but genuine increments. We fix $z = 1$ rather than adapt it to the local signal-to-noise ratio, and the choice is deliberate. A single, unit threshold is model-free and needs no per-point estimate of the true signal, so it transfers unchanged to a hardware experiment where no exact reference exists, which is the setting the estimator is built for. An adaptive z tuned to the local signal-to-noise ratio could recover some of the small genuine increments lost near the ideal limit, but it would require a noise or signal model that the hardware setting does not supply, and it risks reintroducing the very rectification bias the threshold removes; we therefore keep the fixed, portable choice and report its residual over-correction openly. The near-unbiasedness is not a fine-grid artifact: on a coarser grid ($\delta t = 0.3$) the debiased bias stays small (its magnitude below 0.03), and the sign pattern is stable under the threshold z, which we sweep over $\{0.5, 1, 2\}$ in Appendix A. Readout mitigation does not address this bias at all, because it corrects the measurement confusion matrix and not the rectification. The estimator to trust therefore depends on the regime: the raw estimate where the signal is strong, the debiased estimate where noise dominates.

5.2 What the confidence interval means

We construct a bootstrap confidence interval by resampling the shot counts [50, 51]. It is tempting to read this interval as a statement about the true value. It is not. The bootstrap captures the sampling spread of the estimator around its own expectation, and the estimator is biased, so the interval need not contain the truth. We measure its coverage directly over 50 seeds: the 95% interval covers the exact noisy value in 100%, 20%, 4%, and 0% of cases as the noise scale increases, with Wilson intervals on these coverage figures of $[0.93, 1.00]$, $[0.11, 0.33]$, $[0.01, 0.13]$, and $[0.00, 0.07]$. The interval reports precision, not accuracy. We therefore adopt a binding convention: the confidence interval is always presented as the precision of a named estimator,

and accuracy is reported separately through the bias curves. A value written as a number plus or minus an interval, without a named estimator and a stated grid, is not reported.

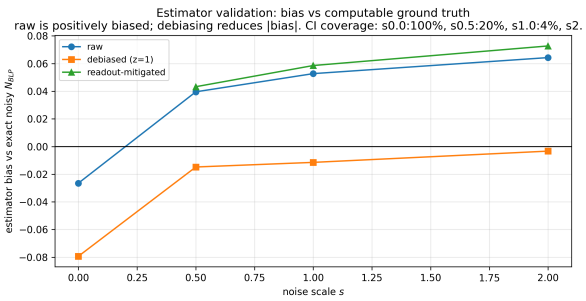
5.3 A hardware-portable control

Hardware has no computable ground truth, so the bias must be estimated some other way. We carry a Markovian null floor: the apparent BLP value of a Markovian control, whose true value is zero, measured on the same grid, shots, and noise,

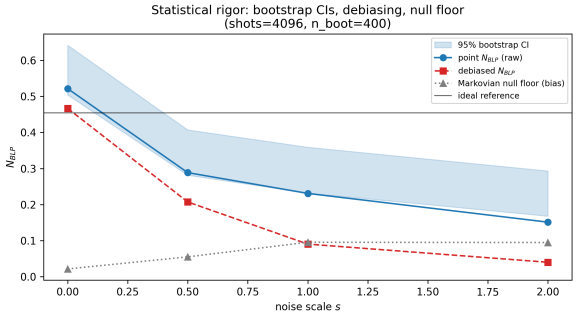
$$\mathcal{N}_{\mathrm{BLP}}^{\mathrm{null}} = \mathcal{N}_{\mathrm{BLP}}\big[\text{Markovian control}\big]\Big|_{\text{same grid, shots, noise}}. \tag{17}$$

Its full-resolution values track the rectification bias closely, rising from 0.022 at zero noise to 0.096 at scale one. The decision rule is simple: a memory claim requires the debiased estimate to exceed the measured null floor. Because the null floor uses only a control experiment and no exact reference, it transfers directly to a hardware setting, where these controls are currently absent. Figure 6 shows the estimator bias and the bootstrap picture together.

One further property must be stated. The discrete measure of Eq. (11) depends on the time grid, because a denser grid rectifies more noise. On a coarse grid the zero-noise raw estimate falls below the true value; on a fine grid it rises above the ideal value. We therefore compare BLP values only within a fixed grid, and every reported value names its grid.



(a) Estimator bias against the exact noisy value. The raw estimator is biased low near the ideal limit and high under noise; the debiased estimator is near-unbiased where noise dominates.



(b) Bootstrap interval, debiased estimate, and Markovian null floor. The interval reports precision; the null floor tracks the bias and ports to hardware.

Figure 6: Finite-shot estimation of the BLP measure, validated against computable ground truth. Under realistic noise the raw estimate can nearly double the true value; the debiased estimator and the null-floor control restore an honest reading.

6 Device noise

We now ask how device noise affects the measured memory, and in particular whether the answer depends on the temporal structure of that noise. The question is sharp for a non-Markovian study, because real device noise is itself correlated in time, while the standard calibrated noise models are memoryless [40].

6.1 Washout and partial recovery

We first apply a scaled depolarizing and readout model. As the noise scale increases from zero to two, the measured $\mathcal{N}_{\mathrm{BLP}}$ falls monotonically from 0.521 to 0.151 on the reference grid. Readout mitigation recovers a small amount. Zero-noise extrapolation, folding the two-qubit gates to

amplify their noise and extrapolating to the zero-noise limit [52–55], recovers more: at scale one it returns 0.311 against a raw value of 0.231. We use linear (order-one) extrapolation on the fold scales $\{1, 3, 5\}$, the more noise-tolerant variant. The extrapolation amplifies only the two-qubit gate noise, so it leaves a known residual from single-qubit and readout error, which we state rather than hide. A model calibrated to a real device (FakeManilaV2) returns a measured $\mathcal{N}_{\mathrm{BLP}}$ of exactly zero at the finer $\delta t = 0.05$ grid used for the noise-model comparison, consistent with complete decoherence of the signature; for scale, the pseudomode circuit at the $\delta t = 0.1$ grid already uses about 630 two-qubit gates (Section 7). This quantifies the hardware gap: the circuits that reproduce the memory faithfully on a simulator are, at present depths, beyond a real device [39].

The gap does not, however, put the controls of Section 5 out of reach. An experimentalist can test them on a deliberately scaled-down instance that trades physical reach for depth. Shortening the total evolution to the first one or two revivals and coarsening the Trotter step reduces the number of steps, and with it the two-qubit gate count, into a range that current hardware can run; the price is that fewer revivals are resolved and the absolute BLP value is smaller, not that the diagnostic changes. On such an instance the Markovian null floor of Eq. (17) is directly deployable, because it is a control experiment rather than a computed reference: one runs the Markovian circuit at the same grid, shots, and device, reads off its apparent BLP value as the in-situ rectification floor, and admits a memory claim only when the debiased estimate on the non-Markovian circuit exceeds it. This calibrates the finite-shot bias on the device itself, with no exact reference, and is the concrete first hardware test we recommend before attempting the full-depth circuits.

6.2 Correlated noise is conflated with signal

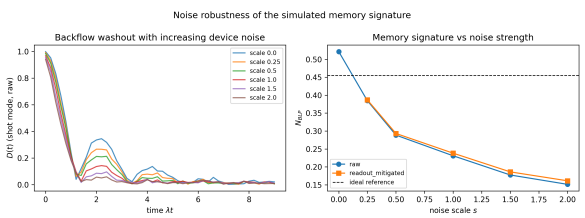
The central noise result concerns temporal structure. We compare two dephasing models matched on their accumulated variance σ_{tot}^2. One is quasi-static: a single random detuning held constant across the evolution, which is correlated in time. The other is memoryless: an independent random detuning at each step. We average over an ensemble of 60 realizations and bootstrap over ensemble members for confidence intervals. Table 4 shows the result. The correlated noise preserves the measured memory significantly more than the memoryless noise, and the gap grows with strength, from 0.005 at $\sigma_{\mathrm{tot}} = 0.3$ to 0.056 at $\sigma_{\mathrm{tot}} = 1.2$. The confidence intervals separate at every strength. Figure 7 shows the washout and the conflation together.

Table 4: Correlated versus memoryless dephasing at matched accumulated strength. Values are ensemble means; the difference is significant, with separated bootstrap 95% intervals at every σ_{tot}.

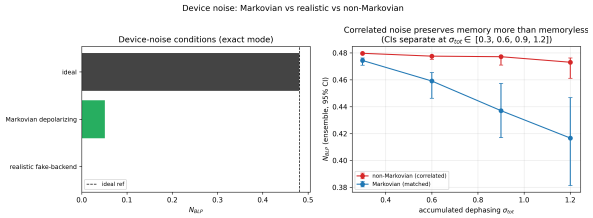
σ_{tot}	correlated $\mathcal{N}_{\mathrm{BLP}}$	memoryless $\mathcal{N}_{\mathrm{BLP}}$	difference
0.3	0.4797	0.4744	+0.0053
0.6	0.4776	0.4591	+0.0185
0.9	0.4772	0.4370	+0.0401
1.2	0.4731	0.4167	+0.0564

The interpretation is a caution. The BLP measure cannot separate physical memory that survives the noise from memory-like temporal correlation that the noise itself carries. Correlated device noise therefore contaminates the diagnostic in the direction of reporting more memory, not less. A study that models device noise as memoryless will misjudge this contribution. Stating the effect inside the simulation, where the two contributions can be controlled independently, answers the objection that a calibrated Markovian model would otherwise leave open. The quasi-static dephasing is applied stroboscopically, one coherent kick per Trotter step, so the

small-σ_{tot} gap could in principle be a discretization artifact. It is not: halving the step from $\delta t = 0.1$ to $\delta t = 0.05$ changes the gap by at most 0.004 at $\sigma_{\mathrm{tot}} \in \{0.3, 0.6\}$, and the intervals separate at both grids (Appendix A).



(a) Washout of the measured memory with noise, and partial recovery by readout mitigation and zero-noise extrapolation.



(b) Correlated versus memoryless dephasing at matched strength. The correlated model preserves more memory, with separated confidence intervals.

Figure 7: Device noise. Memoryless noise washes the signature out (a), while temporally correlated noise is conflated with physical memory (b).

7 A resource-fair embedding comparison

A natural question is whether the pseudomode is the most economical embedding. An earlier version of this comparison used a collision model restricted to a single reused memory ancilla. That model cannot reproduce a Lorentzian backflow, and it loses to the pseudomode. The comparison was unfair, and we retract the conclusion drawn from it. We report the retraction openly, because the corrected comparison is the informative one.

A faithful finite-memory collision model discretizes the bath into a chain of ancillas with a lossy boundary, in the spirit of a chain mapping [32, 35, 36]. The system injects at the first site, excitation hops along the chain, reflects at the finite end, and returns, which produces a tunable memory time. Weak damping at the far end supplies dissipation. The qubit count grows linearly with the chain length M, giving $M + 2$ qubits including one damping ancilla. We implement this as a circuit and verify that it reproduces the state-vector engine to 10^{-13}, so the resource and noise claims rest on the same map.

We fit each chain length to the same Lorentzian population target and evaluate on the independent BLP measure. This separation of the fit target from the evaluation observable guards against circularity. Table 5 reports the result. The pseudomode uses three qubits and 630 two-qubit gates and reaches a BLP error of 0.0022. A chain of length two uses four qubits and reaches an error of 0.0003. The two embeddings are at parity at equal qubit cost for the single-Lorentzian bath. Their noise robustness is also at parity: under the depolarizing model at scale one, the pseudomode returns 0.231 and the chain returns 0.219, a difference within the shot-level scatter.

Table 5: Resource-fair comparison for the single-Lorentzian bath. The chain uses a global optimizer for its fit; all engines agree to 10^{-13}. Two-qubit gate count grows linearly with chain length.

embedding	qubits	CX count	$\mathcal{N}_{\mathrm{BLP}}$ error
pseudomode	3	630	0.0022
chain, M=1	3	630	0.0033
chain, M=2	4	990	0.0003
chain, M=3	5	1350	0.0320

The physical reason for the parity is transparent. At zero temperature the single-Lorentzian bath needs exactly one damped mode, and one damped chain qubit is that mode. The embeddings should diverge only where the Fock rule of Section 4 demands more than two levels, since a mode holds extra levels cheaply while a qubit chain must add ancillas. We state this as the natural continuation rather than demonstrate it here. Figure 8 shows the accuracy against qubit cost.

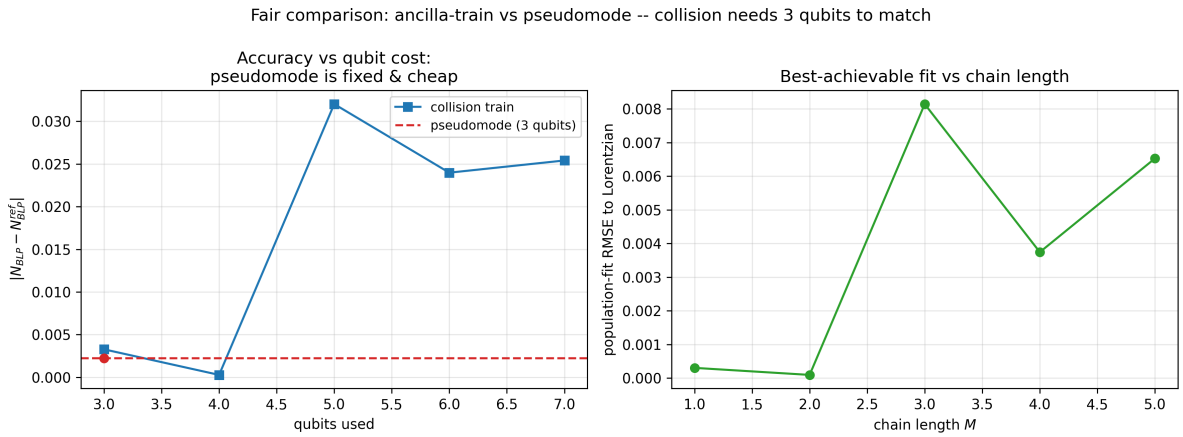


Figure 8: Resource-fair embedding comparison for the single-Lorentzian bath. The pseudomode (three qubits) and the ancilla-train collision model reach parity in accuracy at equal qubit cost, and their fit quality and noise robustness track each other. The retracted single-ancilla model is not shown.

8 Generalization: an excitonic dimer

To show that the pipeline extends beyond a single system qubit, we simulate a two-site excitonic dimer, the minimal unit of environment-assisted energy transport [56, 57]. The system Hamiltonian is

$$H_S = \varepsilon_L\, n_L + \varepsilon_R\, n_R + J\,(\sigma_L^+\sigma_R^- + \sigma_L^-\sigma_R^+), \tag{18}$$

with a site-energy detuning $\varepsilon_{L,R} = \pm 0.5$ and an excitonic coupling $J = 1$. Each site is dephased by its own pseudomode. The circuit uses five qubits. It reproduces the QuTiP site populations to a maximum deviation of 0.024 against a four-level reference, where the residual is dominated by the expected two-level truncation. The four-level reference is itself not fully converged, so this 0.024 is a lower bound on the true truncation error; the matched-truncation deviation, quoted at a finer grid, is a pure Trotter error of 0.0013. The transfer efficiency to the acceptor site is 0.772, and the acceptor population shows coherent revivals (supplementary Figure 10, Appendix D). The dimer therefore inherits both the framework and its truncation caveat: even a physically motivated model needs more than two Fock levels for a faithful memory value.

9 Discussion and limitations

The results share a single message. If a simulation is used to measure memory, the measurement must be shown to be faithful under the conditions of the simulation. We have shown three ways it fails. Truncation misrepresents the measure with a regime-dependent sign, and a variance-aware rule states the dimension needed to avoid it. Finite sampling biases the measure by up to the size of the signal, and a debiased estimator with a null-floor control restores an honest reading. Correlated device noise is conflated with physical memory, which a memoryless noise model cannot capture. The controls we provide, the debiased estimator of Eq. (16), the null floor of

Eq. (17), and the fixed-grid convention, are designed to transfer to a hardware experiment, which is where they are most needed and currently absent [34, 37].

The limitations are explicit. The work runs on classical simulators; the calibrated-noise result quantifies the hardware gap but does not close it. The baths are of Lorentzian type, representable by pseudomodes, and we do not treat a continuum Ohmic bath; a chain mapping or the thermofield approach would extend the reach [35, 58]. We use the BLP measure as the primary quantity; we also compute the RHP measure, which corroborates the finite-temperature destruction but is numerically unstable at strong coupling (Section 3.4), so it serves as a qualitative cross-check rather than a precise second measure. The BLP measure is a discrete sum and depends on the grid, which we handle by a fixed-grid convention. The Fock rule is empirical over its stated domain rather than a theorem, and its constant must be calibrated on super-Poissonian statistics. Outside Stage A, absolute BLP values from a coarse canonical pair are lower bounds, certified against a Bloch-sphere optimization: the largest absolute gap is 0.075, which is 24% in relative terms, at $g = 0.4$, while relative gaps reach 53% only where the absolute value is negligible ($g = 0.2$).

The natural continuations are the regime where the two embeddings diverge, set by the Fock rule; the extension to finite-temperature and continuum baths; and a hardware demonstration of the null-floor control, which does not require an exact reference.

10 Conclusion

We have presented a validated, simulator-based framework for quantifying non-Markovian memory in digital quantum simulation, and we have used it to show where the quantification breaks. A single-qubit mode encoding misrepresents the BLP measure by about a factor of 2.6 at strong coupling and destroys it at moderate temperature. The required Fock dimension follows the occupied number extent through a cross-validated rule. The finite-shot estimate of the measure can nearly double the true value under realistic noise, and a debiased estimator with a hardware-portable null floor corrects it. Correlated device noise is conflated with physical memory by the measure. A resource-fair comparison places the pseudomode and an ancilla-train collision model at parity for a single-Lorentzian bath. All results are reproducible from recorded data, and none of them claims a quantum advantage. The contribution is a set of diagnostics and controls that make simulated, and eventually measured, non-Markovianity honest.

Data availability statement

The data that support the findings of this study are openly available. All quantitative results and figures are generated from recorded JSON files with fixed random seeds; the code, tests, and data records are available at `https://github.com/tom32bit/Benchmarking-Non-Markovianity-Measures-in-Digital-Quantum-Simulation-of-Open-Quantum-Systems` and archived with a permanent DOI at Zenodo (DOI to be inserted on acceptance).

Conflict of interest

The author declares no conflict of interest.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Acknowledgments

The author thanks the maintainers of QuTiP and Qiskit.

Declaration of the use of AI tools

S.M.Y.I.T. conceived and designed the study, implemented the code, performed the analysis, and wrote the manuscript. AI-based assistance tools were used during code development and manuscript preparation. The author verified every result against recorded data and takes full responsibility for the content. The disclosure wording will be aligned with the journal's current policy on the use of AI tools at submission.

A Statistical robustness

Table 6 reports the debiased-estimator bias under the three thresholds $z \in \{0.5, 1, 2\}$ together with the bootstrap-CI coverage of the exact noisy value and a Wilson 95% interval on that coverage, over 50 seeds on the $\delta t = 0.15$ grid. The debiased bias is small and its sign pattern is stable across z; the coverage is well below the nominal 95% once bias is present, which is why the interval is reported as a precision, not an accuracy.

Table 6: Estimator statistical robustness (50 seeds, 4096 shots, $\delta t = 0.15$). Debiased bias at three thresholds, and bootstrap-CI coverage of the truth with a Wilson interval.

scale	deb. bias (z=0.5)	(z=1)	(z=2)	CI coverage	Wilson 95%
0.0	−0.064	−0.079	−0.105	1.00	[0.93, 1.00]
0.5	+0.008	−0.015	−0.056	0.20	[0.11, 0.33]
1.0	+0.015	−0.012	−0.049	0.04	[0.01, 0.13]
2.0	+0.019	−0.003	−0.025	0.00	[0.00, 0.07]

Table 7 shows the grid robustness. On a coarser grid ($\delta t = 0.3$, 20 seeds) the debiased bias stays small, so the near-unbiasedness of Section 5.1 is not a fine-grid artifact, although its precise value is grid-dependent. The correlated-noise conflation gap of Section 6.2 changes by at most 0.004 between $\delta t = 0.1$ and $\delta t = 0.05$, and the bootstrap intervals separate at both grids, so the small-σ_{tot} conflation is not a stroboscopic artifact.

Table 7: Grid robustness. Left: coarse-grid estimator bias. Right: conflation gap under a halved Trotter step, with interval separation at both grids.

scale	true	raw	debiased	σ_{tot}	gap (δt=0.1)	gap (δt=0.05)	sep.
0.0	0.382	+0.054	+0.037	0.3	+0.0022	+0.0031	both
0.5	0.303	−0.002	−0.028	0.6	+0.0082	+0.0121	both
1.0	0.239	+0.001	−0.024				
2.0	0.146	+0.010	−0.020				

B Fock-rule validation domain

The Fock rule of Eq. (14) is validated over the following domain, outside which it is an extrapolation rather than a bound. The temperature family uses $\bar{n} \in \{0.1, 0.25, 0.5, 0.75, 1.0, 1.5, 2.0\}$ with reference dimension 10; the coupling family uses $g \in \{0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8\}$ with reference

dimension 8; both use $t_{\max} = 8$, $\delta t = 0.05$, and the accuracy targets $\epsilon \in \{3 \times 10^{-2}, 1 \times 10^{-2}, 3 \times 10^{-3}\}$. The safe constant is $c = 3$. Table 8 gives the required dimension $d_{\mathrm{req}}(\epsilon)$ at each point; the single right-censored entry ($\bar{n} = 2.0$ at the strictest target) is marked. The Bloch-sphere pair optimization of Section 2.3 uses a fixed budget: the three canonical antipodal axes as seeds, 40 random antipodal starts, and 30 coordinate-ascent refinement steps, with random seed 1, so the certified gaps are search-converged rather than search-limited.

Table 8: Required Fock dimension $d_{\mathrm{req}}(\epsilon)$ across the validation domain, at the three accuracy targets. "c" marks a right-censored entry that reaches the reference dimension.

family	param	3×10^{-2}	1×10^{-2}	3×10^{-3}	family	param	3×10^{-2}	1×10^{-2}	3×10^{-3}
temp.	0.10	2	3	4	coup.	0.2	2	2	3
temp.	0.25	3	3	5	coup.	0.3	3	3	4
temp.	0.50	3	4	6	coup.	0.4	3	4	4
temp.	0.75	4	5	7	coup.	0.5	3	4	5
temp.	1.00	4	6	8	coup.	0.6	3	4	5
temp.	1.50	5	8	9	coup.	0.7	3	5	6
temp.	2.00	6	8	10c	coup.	0.8	4	5	6

Table 9 makes the cross-validation of the constant explicit. For each of $c \in \{2.0, 2.75, 3.0\}$ it reports, per family, whether the rule of Eq. (14) is a safe upper bound on all seven points, the largest over-provision (extra Fock levels beyond d_{req}), and the number of points it under-provisions. The pattern is the one stated in Section 4: $c = 2.0$, the value the coupling family alone would justify, is safe on coupling but under-provisions three of the seven thermal points by up to two levels; $c = 2.75$ is the smallest constant that is safe on both; and $c = 3.0$ keeps a one-level margin on the thermal family at the cost of at most three surplus levels on the coupling family. The d_{req} values here are taken at the headline target $\epsilon = 1 \times 10^{-2}$.

Table 9: Over- and under-provisioning of the variance-aware Fock rule across the constant c, evaluated separately on the seven temperature and seven coupling points at $\epsilon = 1 \times 10^{-2}$. "max over" is the largest surplus of rule dimension over d_{req}; "# under" counts points where the rule falls below d_{req} (a value of zero means the rule is a safe bound on that family).

	temperature family			coupling family		
c	safe	max over	# under	safe	max over	# under
2.00	no	0	3	yes	1	0
2.75	yes	1	0	yes	2	0
3.00	yes	1	0	yes	3	0

C RHP values for the truncation study

Table 10 lists the RHP values that underlie Section 3.4. At finite temperature the circuit value lies below the converged value at every point, corroborating the BLP destruction, though the estimator is ill-conditioned (the ill-conditioned fraction rises to 0.53 near the $|G(t)| = 0$ crossings). At strong coupling no intermediate map is flagged ill-conditioned, yet the converged value is non-monotonic in coupling and the circuit-converged ordering reverses at $g = 0.7$ relative to BLP, so we treat the strong-coupling RHP as numerically unstable. Values are rounded; the measure is large and noisy, and only the qualitative ordering is meaningful.

Table 10: RHP measure for the two-level circuit and the converged reference. Left: finite temperature, with the ill-conditioned fraction. Right: strong coupling (all ill-conditioned fractions are zero).

$\bar{n}$	circ.	conv.	ill. (circ.)	ill. (conv.)	g	circ.	conv.
0.00	204	204	0.04	0.04	0.3	16	23
0.10	75	157	0.06	0.01	0.4	10	74
0.25	88	113	0.23	0.03	0.5	39	17
0.50	83	308	0.36	0.14	0.7	42	143
0.75	61	121	0.47	0.20			
1.00	43	128	0.53	0.21			
1.50	7	112	0.07	0.23			

D Supplementary figures

The main text summarizes the truncation study through the tables and the phase diagram (Figure 4). Figure 9 collects the scaling laws referenced in Section 2.4, and Figure 10 shows the excitonic-dimer energy transfer of Section 8.

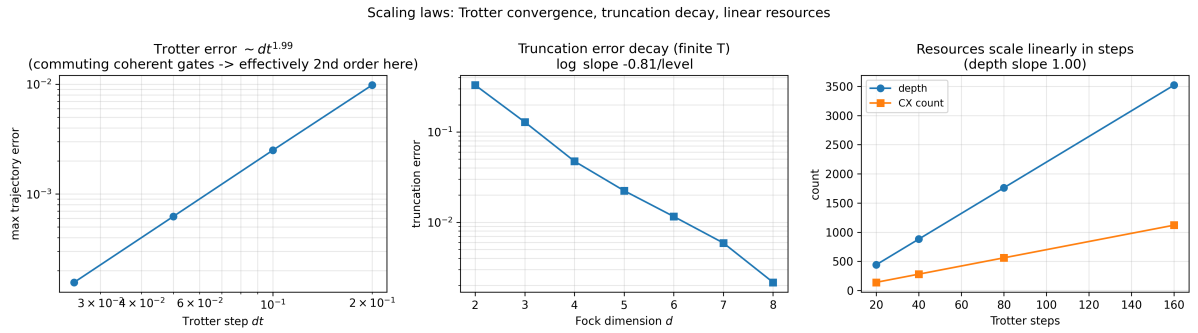


Figure 9: Scaling laws. Left: the Trotter error against step size, with a fitted order of 1.99; the coherent gates commute, so the Hamiltonian part is exact and the leading error is the effectively second-order coherent-dissipative split. Middle: the Fock-truncation error decays roughly geometrically with dimension, about a factor of two per level. Right: the transpiled circuit depth and two-qubit gate count grow linearly in the number of Trotter steps.

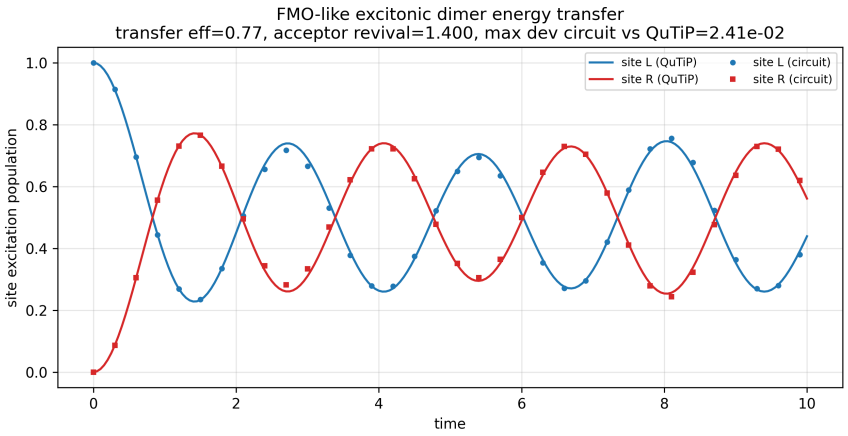


Figure 10: Excitonic dimer energy transfer. Site populations for the donor and acceptor from the five-qubit circuit (markers) against the QuTiP reference (lines). The excitation transfers coherently between sites and shows revivals, a signature of the structured environment. The circuit reproduces the reference to a maximum deviation of 0.024, dominated by the two-level mode truncation.

References

[1] Heinz-Peter Breuer and Francesco Petruccione. *The Theory of Open Quantum Systems*. Oxford University Press, Oxford, 2002.

[2] Ulrich Weiss. *Quantum Dissipative Systems*. World Scientific, 4 edition, 2012.

[3] Heinz-Peter Breuer, Elsi-Mari Laine, Jyrki Piilo, and Bassano Vacchini. Colloquium: Non-markovian dynamics in open quantum systems. *Rev. Mod. Phys.*, 88:021002, 2016. doi: 10.1103/RevModPhys.88.021002.

[4] Inés de Vega and Daniel Alonso. Dynamics of non-markovian open quantum systems. *Rev. Mod. Phys.*, 89:015001, 2017. doi: 10.1103/RevModPhys.89.015001.

[5] Ángel Rivas, Susana F. Huelga, and Martin B. Plenio. Quantum non-markovianity: characterization, quantification and detection. *Rep. Prog. Phys.*, 77:094001, 2014. doi: 10.1088/0034-4885/77/9/094001.

[6] A. J. Leggett, S. Chakravarty, A. T. Dorsey, Matthew P. A. Fisher, Anupam Garg, and W. Zwerger. Dynamics of the dissipative two-state system. *Rev. Mod. Phys.*, 59:1, 1987. doi: 10.1103/RevModPhys.59.1.

[7] Akihito Ishizaki and Graham R. Fleming. Theoretical examination of quantum coherence in a photosynthetic system at physiological temperature. *Proc. Natl. Acad. Sci. U.S.A.*, 106: 17255, 2009. doi: 10.1073/pnas.0908989106.

[8] Heinz-Peter Breuer, Elsi-Mari Laine, and Jyrki Piilo. Measure for the degree of non-markovian behavior of quantum processes in open systems. *Phys. Rev. Lett.*, 103:210401, 2009. doi: 10.1103/PhysRevLett.103.210401.

[9] Elsi-Mari Laine, Jyrki Piilo, and Heinz-Peter Breuer. Measure for the non-markovianity of quantum processes. *Phys. Rev. A*, 81:062115, 2010. doi: 10.1103/PhysRevA.81.062115.

[10] Ángel Rivas, Susana F. Huelga, and Martin B. Plenio. Entanglement and non-markovianity of quantum evolutions. *Phys. Rev. Lett.*, 105:050403, 2010. doi: 10.1103/PhysRevLett.105.050403.

[11] Shunlong Luo, Shuangshuang Fu, and Hongting Song. Quantifying non-markovianity via correlations. *Phys. Rev. A*, 86:044101, 2012. doi: 10.1103/PhysRevA.86.044101.

[12] Dariusz Chruściński and Sabrina Maniscalco. Degree of non-markovianity of quantum evolution. *Phys. Rev. Lett.*, 112:120404, 2014. doi: 10.1103/PhysRevLett.112.120404.

[13] M. M. Wolf, J. Eisert, T. S. Cubitt, and J. I. Cirac. Assessing non-markovian quantum dynamics. *Phys. Rev. Lett.*, 101:150402, 2008. doi: 10.1103/PhysRevLett.101.150402.

[14] Li Li, Michael J. W. Hall, and Howard M. Wiseman. Concepts of quantum non-markovianity: A hierarchy. *Phys. Rep.*, 759:1, 2018. doi: 10.1016/j.physrep.2018.07.001.

[15] Simon Milz and Kavan Modi. Quantum stochastic processes and quantum non-markovian phenomena. *PRX Quantum*, 2:030201, 2021. doi: 10.1103/PRXQuantum.2.030201.

[16] Felix A. Pollock, César Rodríguez-Rosario, Thomas Frauenheim, Mauro Paternostro, and Kavan Modi. Non-markovian quantum processes: Complete framework and efficient characterization. *Phys. Rev. A*, 97:012127, 2018. doi: 10.1103/PhysRevA.97.012127.

[17] Bi-Heng Liu, Li Li, Yun-Feng Huang, Chuan-Feng Li, Guang-Can Guo, Elsi-Mari Laine, Heinz-Peter Breuer, and Jyrki Piilo. Experimental control of the transition from markovian to non-markovian dynamics of open quantum systems. *Nat. Phys.*, 7:931, 2011. doi: 10.1038/nphys2085.

[18] Göran Lindblad. On the generators of quantum dynamical semigroups. *Commun. Math. Phys.*, 48:119, 1976. doi: 10.1007/BF01608499.

[19] Vittorio Gorini, Andrzej Kossakowski, and E. C. G. Sudarshan. Completely positive dynamical semigroups of N-level systems. *J. Math. Phys.*, 17:821, 1976. doi: 10.1063/1.522979.

[20] I. M. Georgescu, S. Ashhab, and Franco Nori. Quantum simulation. *Rev. Mod. Phys.*, 86: 153, 2014. doi: 10.1103/RevModPhys.86.153.

[21] B. M. Garraway. Nonperturbative decay of an atomic system in a cavity. *Phys. Rev. A*, 55: 2290, 1997. doi: 10.1103/PhysRevA.55.2290.

[22] Dario Tamascelli, Andrea Smirne, Susana F. Huelga, and Martin B. Plenio. Nonperturbative treatment of non-markovian dynamics of open quantum systems. *Phys. Rev. Lett.*, 120: 030402, 2018. doi: 10.1103/PhysRevLett.120.030402.

[23] F. Mascherpa, A. Smirne, A. D. Somoza, P. Fernández-Acebal, S. Donadi, D. Tamascelli, S. F. Huelga, and M. B. Plenio. Optimized auxiliary oscillators for the simulation of general open quantum systems. *Phys. Rev. A*, 101:052108, 2020. doi: 10.1103/PhysRevA.101.052108.

[24] L. K. Zhou, G. R. Jin, and W. Yang. Systematic and efficient pseudomode method to simulate open quantum systems under a bosonic environment. *Phys. Rev. A*, 110:022221, 2024. doi: 10.1103/PhysRevA.110.022221.

[25] Yoshitaka Tanimura and Ryogo Kubo. Time evolution of a quantum system in contact with a nearly gaussian-markoffian noise bath. *J. Phys. Soc. Jpn.*, 58:101, 1989. doi: 10.1143/JPSJ.58.101.

[26] Yoshitaka Tanimura. Numerically "exact" approach to open quantum dynamics: The hierarchical equations of motion (heom). *J. Chem. Phys.*, 153:020901, 2020. doi: 10.1063/5.0011599.

[27] Neill Lambert, Tarun Raheja, Simon Cross, Paul Menczel, Shahnawaz Ahmed, Alexander Pitchford, Daniel Burgarth, and Franco Nori. QuTiP-BoFiN: A bosonic and fermionic numerical hierarchical-equations-of-motion library with applications in light-harvesting, quantum control, and single-molecule electronics. *Phys. Rev. Research*, 5:013181, 2023. doi: 10.1103/PhysRevResearch.5.013181.

[28] Xiaohan Dan, Eitan Geva, and Victor S. Batista. Simulating non-markovian quantum dynamics on NISQ computers using the hierarchical equations of motion. *J. Chem. Theory Comput.*, 2025. doi: 10.1021/acs.jctc.4c01565.

[29] M. Kliesch, T. Barthel, C. Gogolin, M. Kastoryano, and J. Eisert. Dissipative quantum church-turing theorem. *Phys. Rev. Lett.*, 107:120501, 2011. doi: 10.1103/PhysRevLett.107.120501.

[30] Richard Cleve and Chunhao Wang. Efficient quantum algorithms for simulating Lindblad evolution. In *44th International Colloquium on Automata, Languages, and Programming (ICALP 2017)*, 2017. doi: 10.4230/LIPIcs.ICALP.2017.17.

[31] Francesco Ciccarello, G. Massimo Palma, and Vittorio Giovannetti. Collision-model-based approach to non-markovian quantum dynamics. *Phys. Rev. A*, 87:040103(R), 2013. doi: 10.1103/PhysRevA.87.040103.

[32] Francesco Ciccarello, Salvatore Lorenzo, Vittorio Giovannetti, and G. Massimo Palma. Quantum collision models: Open system dynamics from repeated interactions. *Phys. Rep.*, 954:1, 2022. doi: 10.1016/j.physrep.2022.01.001.

[33] Hans Michael Christensen, Johannes Agerskov, and Frederik Nathan. Ancilla-train quantum algorithm for simulating non-markovian open quantum systems. Preprint arXiv:2509.12717 [quant-ph], 2025.

[34] Charlotte Bäcker, Krishna Palaparthy, and Walter T. Strunz. Revealing the quantum nature of memory in non-markovian dynamics on IBM quantum. *New J. Phys.*, 2025. doi: 10.1088/1367-2630/ae5f3d.

[35] Alex W. Chin, Ángel Rivas, Susana F. Huelga, and Martin B. Plenio. Exact mapping between system-reservoir quantum models and semi-infinite discrete chains using orthogonal polynomials. *J. Math. Phys.*, 51:092109, 2010. doi: 10.1063/1.3490188.

[36] Javier Prior, Alex W. Chin, Susana F. Huelga, and Martin B. Plenio. Efficient simulation of strong system-environment interactions. *Phys. Rev. Lett.*, 105:050404, 2010. doi: 10.1103/PhysRevLett.105.050404.

[37] Kade Head-Marsden, Stefan Krastanov, David A. Mazziotti, and Prineha Narang. Capturing non-markovian dynamics on near-term quantum computers. *Phys. Rev. Research*, 3:013182, 2021. doi: 10.1103/PhysRevResearch.3.013182.

[38] Anthony W. Schlimgen, Kade Head-Marsden, LeeAnn M. Sager, Prineha Narang, and David A. Mazziotti. Quantum simulation of open quantum systems using a unitary decomposition of operators. *Phys. Rev. Lett.*, 127:270503, 2021. doi: 10.1103/PhysRevLett.127.270503.

[39] John Preskill. Quantum computing in the NISQ era and beyond. *Quantum*, 2:79, 2018. doi: 10.22331/q-2018-08-06-79.

[40] Zhenyu Cai, Ryan Babbush, Simon C. Benjamin, Suguru Endo, William J. Huggins, Ying Li, Jarrod R. McClean, and Thomas E. O'Brien. Quantum error mitigation. *Rev. Mod. Phys.*, 95:045005, 2023. doi: 10.1103/RevModPhys.95.045005.

[41] Seth Lloyd. Universal quantum simulators. *Science*, 273:1073, 1996. doi: 10.1126/science.273.5278.1073.

[42] Masuo Suzuki. Generalized trotter's formula and systematic approximants of exponential operators and inner derivations with applications to many-body problems. *Commun. Math. Phys.*, 51:183, 1976. doi: 10.1007/BF01609348.

[43] W. Forrest Stinespring. Positive functions on C*-algebras. *Proc. Amer. Math. Soc.*, 6:211, 1955. doi: 10.1090/S0002-9939-1955-0069403-4.

[44] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 10th anniversary edition, 2010.

[45] Jarrod R. McClean, Sergio Boixo, Vadim N. Smelyanskiy, Ryan Babbush, and Hartmut Neven. Barren plateaus in quantum neural network training landscapes. *Nat. Commun.*, 9:4812, 2018. doi: 10.1038/s41467-018-07090-4.

[46] Steffen Wissmann, Antti Karlsson, Elsi-Mari Laine, Jyrki Piilo, and Heinz-Peter Breuer. Optimal state pairs for non-markovian quantum dynamics. *Phys. Rev. A*, 86:062108, 2012. doi: 10.1103/PhysRevA.86.062108.

[47] J. R. Johansson, P. D. Nation, and Franco Nori. QuTiP: An open-source Python framework for the dynamics of open quantum systems. *Comput. Phys. Commun.*, 183:1760, 2012. doi: 10.1016/j.cpc.2012.02.021.

[48] J. R. Johansson, P. D. Nation, and Franco Nori. QuTiP 2: A Python framework for the dynamics of open quantum systems. *Comput. Phys. Commun.*, 184:1234, 2013. doi: 10.1016/j.cpc.2012.11.019.

[49] Qiskit contributors. Qiskit: An open-source framework for quantum computing, 2023.

[50] B. Efron. Bootstrap methods: Another look at the jackknife. *Ann. Statist.*, 7:1, 1979. doi: 10.1214/aos/1176344552.

[51] Bradley Efron and Robert J. Tibshirani. *An Introduction to the Bootstrap*. Chapman and Hall/CRC, 1994.

[52] Kristan Temme, Sergey Bravyi, and Jay M. Gambetta. Error mitigation for short-depth quantum circuits. *Phys. Rev. Lett.*, 119:180509, 2017. doi: 10.1103/PhysRevLett.119.180509.

[53] Ying Li and Simon C. Benjamin. Efficient variational quantum simulator incorporating active error minimization. *Phys. Rev. X*, 7:021050, 2017. doi: 10.1103/PhysRevX.7.021050.

[54] Suguru Endo, Simon C. Benjamin, and Ying Li. Practical quantum error mitigation for near-future applications. *Phys. Rev. X*, 8:031027, 2018. doi: 10.1103/PhysRevX.8.031027.

[55] Abhinav Kandala, Kristan Temme, Antonio D. Córcoles, Antonio Mezzacapo, Jerry M. Chow, and Jay M. Gambetta. Error mitigation extends the computational reach of a noisy quantum processor. *Nature*, 567:491, 2019. doi: 10.1038/s41586-019-1040-7.

[56] Martin B. Plenio and Susana F. Huelga. Dephasing-assisted transport: quantum networks and biomolecules. *New J. Phys.*, 10:113019, 2008. doi: 10.1088/1367-2630/10/11/113019.

[57] Patrick Rebentrost, Masoud Mohseni, Ivan Kassal, Seth Lloyd, and Alán Aspuru-Guzik. Environment-assisted quantum transport. *New J. Phys.*, 11:033003, 2009. doi: 10.1088/1367-2630/11/3/033003.

[58] Dario Tamascelli, Andrea Smirne, James Lim, Susana F. Huelga, and Martin B. Plenio. Efficient simulation of finite-temperature open quantum systems. *Phys. Rev. Lett.*, 123: 090402, 2019. doi: 10.1103/PhysRevLett.123.090402.