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.

Spectral Decay and Fading-Memory Bounds for Reset-Gated Quantum Recurrent Maps

Parag Vijay Pawar
pawarpv23.meta@coeptech.ac.in
COEP Technological University https://orcid.org/0009-0006-5191-1455
Pranav Devkar
COEP Technological University https://orcid.org/0009-0007-4284-0445
Rutuja Gabhane
COEP Technological University
Riya Ambekar
COEP Technological University
Mukta Nivelkar
mmn.comp@coeptech.ac.in
COEP Technological University https://orcid.org/0000-0001-7452-6978
Sunil Bhirud
COEP Technological University

Research Article

Keywords: quantum channel, fading memory, stochastic resetting, Fourier spectrum, Lipschitz bound, data re-uploading

Posted Date: November 12th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11690363/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.

Spectral decay and weight-independent fading-memory bounds for reset-gated quantum recurrent maps

Parag Pawar[1*], Pranav Devkar[2], Rutuja Gabhane[3], Riya Ambekar[4], Mukta Nivelkar[5*], Sunil Bhirud[6]

[1–4]Department of Metallurgy and Materials Engineering.
[5,6]Department of Computer Science and Engineering.
COEP Technological University, Pune, 411005, Maharashtra, India.

*Corresponding author(s). E-mail(s): pawarpv23.meta@coeptech.ac.in; mmn.comp@coeptech.ac.in;
Contributing authors: pranavud23.meta@coeptech.ac.in; gabhanerd23.meta@coeptech.ac.in; riyasa23.meta@coeptech.ac.in; vc@coeptech.ac.in;

Abstract

This paper studies a discrete-time recurrent map on density matrices in which the memory state is updated by a data-dependent re-uploading unitary and gated by an input-dependent replacement channel $\mathcal{R}_f(\rho) = f\rho + (1-f)\sigma$, a quantum analogue of the forget gate of a long short-term memory cell. The starting point is the last-renewal representation familiar from stochastic resetting, extended from a fixed unitary and constant reset rate to input-dependent unitaries and input-dependent gates: the state after T steps is a convex combination of T reset-free horizon states with weights in closed form. On this basis three quantitative results are proved. First, as a function of any single input the output is a trigonometric polynomial whose frequencies lie in an explicit finite set determined by the encoding alone, and, for a constant gate, whose coefficients at nonzero frequencies are bounded by $f^m\|O\|$, where m is the age of that input. Second, the sensitivity of the output to each input admits a bound that does not depend on the trainable unitaries, includes the contribution of the gate derivative, and is summable uniformly in T; no bound of this type holds for the classical cell. Third, the representation yields an exactly unbiased estimator of the output, using only reset-free unitary circuits, whose expected depth equals $\sum_m P_m \le (1 - f_{\max})^{-1}$ uniformly in T, together with an explicit truncation bound. All identities are verified numerically to machine precision, and all bounds on random instances, in two independent implementations, and the reconstruction is executed on two superconducting quantum processors.

Keywords: quantum channel, fading memory, stochastic resetting, Fourier spectrum, Lipschitz bound, data re-uploading

1 Introduction

Let $\mathcal{M}_d$ denote the complex $d \times d$ matrices and $\mathcal{S}_d \subset \mathcal{M}_d$ the set of density matrices, that is, the positive semidefinite matrices of unit trace. A *quantum channel* is a completely positive trace-preserving linear map $\mathcal{M}_d \to \mathcal{M}_d$. This paper concerns a particular discrete-time recurrent map built from two families of channels: conjugation

by a data-dependent unitary, and the *replacement channel*

$$\mathcal{R}_f(\rho) = f\rho + (1-f)\sigma, \qquad f \in [0,1], \tag{1.1}$$

which with probability $1-f$ discards the state and replaces it by a fixed reference state $\sigma \in \mathcal{S}_d$.

The motivation comes from recurrent models in machine learning. A classical long short-term memory (LSTM) cell [12] maintains a vector $c_t \in \mathbb{R}^H$ and updates it by $c_t = f_t \odot c_{t-1} + i_t \odot g_t$, where $f_t, i_t \in (0,1)^H$ are *gates* computed from the input and g_t is a candidate update. Several authors have proposed quantum analogues in which the gates are produced by variational quantum circuits while the memory itself remains a classical vector [1, 7]. The map studied here instead promotes the memory to a density matrix and realises the forget gate as the channel (1.1). Two structural features follow immediately and are the subject of this paper: the state is automatically normalised, and the convex structure of $\mathcal{R}_f$ permits an exact decomposition of the recurrence into a finite mixture of unitary trajectories.

1.1 The map

Fix $n \in \mathbb{N}$, write $d = 2^n$, and let X_q denote the Pauli X operator acting on the q-th tensor factor of $(\mathbb{C}^2)^{\otimes n}$. Fix *encoding scales* $\lambda = (\lambda_1, \dots, \lambda_n) \in (0,\infty)^n$ and set

$$\begin{aligned} G &= \sum_{q=1}^{n} \lambda_q X_q, \\ S(x) &= \exp\!\big(-\tfrac{i}{2} x G\big) = \bigotimes_{q=1}^{n} R_X(\lambda_q x), \end{aligned} \tag{1.2}$$

where $R_X(\theta) = \exp(-\frac{i}{2}\theta X)$. Fix $L \in \mathbb{N}$ and unitaries $W_1, \dots, W_L$ on $(\mathbb{C}^2)^{\otimes n}$, which play the role of trainable parameters, and define the *step unitary*

$$U(x) = W_L S(x)\, W_{L-1} S(x) \cdots W_1 S(x). \tag{1.3}$$

The repetition of $S(x)$ across the L layers is the *data re-uploading* construction of [13]. Write $\Phi_x(\rho) = U(x)\rho\, U(x)^*$.

Definition 1.1. Let $\sigma \in \mathcal{S}_d$ be a fixed reference state and let $\varphi\colon \mathbb{R} \to [f_{\min}, f_{\max}]$ with $0 \le f_{\min} \le f_{\max} < 1$ be a C^1 *gate function*. Given an input sequence $x = (x_1, \dots, x_T) \in \mathbb{R}^T$, set $f_t = \varphi(x_t)$ for $t \ge 2$ and adopt throughout the convention

$$\begin{gathered} f_1 := 0 \\ \text{(constant, so that } \varphi' \text{ contributes} \\ \text{nothing at } t = 1). \end{gathered} \tag{1.4}$$

The *reset-gated quantum recurrent map* is defined by

$$\begin{aligned} \rho_1 &= \Phi_{x_1}(\sigma), \\ \rho_t &= \Phi_{x_t}(\mathcal{R}_{f_t}(\rho_{t-1})) \quad (2 \le t \le T). \end{aligned} \tag{1.5}$$

Given a self-adjoint $O \in \mathcal{M}_d$, the *output* is $y_T = \mathrm{Tr}(O\rho_T)$.

Note that (1.4) makes (1.5) consistent: applying $\mathcal{R}_0$ to any state returns σ, so the first line of (1.5) is the $t = 1$ case of the second with ρ_0 arbitrary.

1.2 Results

Throughout, set

$$P_m = \prod_{s=m+1}^{T} f_s \qquad (0 \le m \le T), \tag{1.6}$$

with the convention $P_T = 1$; by (1.4), $P_0 = 0$.

Proposition 3.2 is the renewal representation, which extends a known formula (see Section 1.3): ρ_T is the convex combination $\sum_{k=1}^{T} w_k \rho_{k:T}$, where $\rho_{k:T}$ is obtained by running the unitaries from step k to step T without interruption (a pure state when the reference state is pure), with $w_k = (1 - f_k)P_k$. Theorem 4.1 determines the frequency content of y_T in each input variable and shows that, for a constant gate f, the coefficients at nonzero frequencies have modulus at most $f^m\|O\|$, where m is the age of the input. Theorem 5.1 bounds $|\partial y_T/\partial x_s|$ by a quantity depending only on λ, L, $\|O\|$ and the gate values — never on $W_1, \dots, W_L$ — and shows the total sensitivity is bounded uniformly in T. Corollary 5.4 quantifies the error from truncating the history, and Theorem 6.1 gives an unbiased estimator of y_T whose expected circuit depth is at most $(1-f_{\max})^{-1}$, independently of T.

Section 7 reports a numerical verification of all of these results to machine precision, carried out independently in two implementations, and Section 8 reports an execution of the renewal

representation on two superconducting quantum processors.

1.3 Relation to earlier work

Renewal representation. Proposition 3.2 is not new in its essential idea, and it is stated as a proposition for that reason. Writing a state subject to random resets as an average over the time elapsed since the last reset is the last-renewal equation of the stochastic resetting literature [10]; for discrete-time unitary dynamics interrupted by resets with a constant probability r, Wald, Yao, Platini, Hooley and Carollo [16] write the state as $(1-r)^t U^t|0\rangle\langle 0|U^{-t}+r\sum_{n<t}(1-r)^n U^n|0\rangle\langle 0|U^{-n}$, which is the special case of (3.3) with a fixed unitary and a constant gate. More generally, unravelling a channel written as a convex combination into an ensemble of trajectories is the basis of quantum trajectory methods [4, 9], and the underlying probabilistic argument is a classical renewal argument [11]. The extension to input-dependent unitaries $U(x_t)$ and input-dependent gates $f_t = \varphi(x_t)$ is elementary; its role here is as the tool from which Theorems 4.1, 5.1 and 6.1 are derived.

Reservoirs with resets. In quantum reservoir computing, Chen, Nurdin and Yamamoto [6] consider a dissipative reservoir that is reset to a fixed state with probability ϵ at each step, simulate it on hardware by random selection among unitary circuits and resets, and re-initialise the circuits a finite number M of steps in the past, the error being controlled by geometric contraction in $(1-\epsilon)$. Čindrak, Donvil, Lüdge and Jaurigue [8] restrict a reservoir's memory to a finite window of recent inputs and study the effect empirically. Corollary 5.4 and Theorem 6.1 are close in spirit to these constructions. What Theorem 6.1 adds is that, because the weights are known in closed form, the last-reset time can be sampled classically and the estimator is *exactly* unbiased for the full-history output rather than for a truncated one, with expected depth given by the identity $\sum_m P_m$ and a Hoeffding guarantee. Fading memory and universality for dissipative quantum reservoirs are established in [5, 14]; those results concern the existence of fading-memory maps, whereas Theorem 5.1 gives explicit constants.

Spectrum and Lipschitz bounds. The frequency set Ω of Theorem 4.1 is the frequency set computed by Schuld, Sweke and Meyer [15] for a single re-uploading circuit; what is added here is the bound on the coefficients in terms of the age of an input, which has no counterpart in the non-recurrent case. For non-recurrent models with data re-uploading, Berberich, Fink, Pranjić, Tutschku and Holm [2] prove the Lipschitz bound $2\|M\|\sum_j\|w_j\|\|H_j\|$, which for a fixed encoding is likewise independent of the trainable unitaries; Lemma 2.1 is the corresponding single-step estimate. The content of Theorem 5.1 is the recurrent, gated version: the per-input factor $\prod_{t>s} f_t$, the additional term arising from the input dependence of the gate, and summability uniformly in T.

To the best of the authors' knowledge, and subject to the limits of the literature search carried out, the age-dependent spectral bound of Theorem 4.1, the recurrent weight-independent sensitivity bound of Theorem 5.1, and the exactly unbiased bounded-depth estimator of Theorem 6.1 have not appeared in this form for gated quantum recurrent maps.

Remark 1.2 (Scope)**.** No claim of a quantum computational advantage is made. For the parameters used in Sections 7 and 8 the map is efficiently simulable on a classical computer — indeed that is how the verification is carried out — and Theorem 4.1 shows that on inputs sampled at P points per period the map is contained in a classical trigonometric model with $\lfloor P/2 \rfloor$ harmonics. The results are structural: they describe what this class of maps can and cannot represent, and how accurately its output can be estimated from shallow circuits.

2 Preliminaries

Let $\|A\|$ denote the operator norm and $\|A\|_1 = \operatorname{Tr}\sqrt{A^*A}$ the trace norm. The Hölder-type inequalities

$$\begin{aligned} |\operatorname{Tr}(AB)| &\le \|A\|\|B\|_1, \\ \|ABC\|_1 &\le \|A\|\|B\|_1\|C\|, \end{aligned} \tag{2.1}$$

are used repeatedly, as is the fact that $\|\rho\|_1 = 1$ for $\rho \in \mathcal{S}_d$, so that $\|\rho - \rho'\|_1 \le 2$ for $\rho, \rho' \in \mathcal{S}_d$.

Set

$$\Lambda_1 = \sum_{q=1}^{n} \lambda_q, \qquad \Lambda = L\,\Lambda_1. \tag{2.2}$$

Lemma 2.1. *The map $x \mapsto U(x)$ of (1.3) is real-analytic and $\|U'(x)\| \le \Lambda/2$ for all $x \in \mathbb{R}$.*

Proof. The operators $X_1, \dots, X_n$ act on distinct tensor factors, hence commute, and each has spectrum $\{\pm 1\}$; therefore G of (1.2) has spectrum

$$\operatorname{spec}(G) = \Big\{ \textstyle\sum_{q=1}^{n} s_q \lambda_q : \; s \in \{-1,1\}^n \Big\}, \quad (2.3)$$

and $\|G\| = \Lambda_1$. Since $S(x) = \exp(-\frac{i}{2} x G)$, one has $S'(x) = -\frac{i}{2} G\, S(x)$, whence $\|S'(x)\| = \frac{1}{2}\|G\| = \Lambda_1/2$. Differentiating the product (1.3) by the Leibniz rule gives

$$U'(x) = \sum_{l=1}^{L} \big(W_L S(x) \cdots W_{l+1} S(x)\big)\, W_l\, S'(x) \times \big(W_{l-1} S(x) \cdots W_1 S(x)\big),$$

a sum of L terms each of operator norm at most $\|S'(x)\| = \Lambda_1/2$, because all remaining factors are unitary. Hence $\|U'(x)\| \le L\Lambda_1/2 = \Lambda/2$. □

3 The renewal representation

For $1 \le k \le T$ define the *horizon state*

$$\rho_{k:T} = \big(\Phi_{x_T} \circ \Phi_{x_{T-1}} \circ \cdots \circ \Phi_{x_k}\big)(\sigma), \quad (3.1)$$

obtained by applying the step unitaries from time k to time T to the reference state with no intervening gate. If $\sigma = |0\rangle\langle 0|$ is pure then $\rho_{k:T} = |\psi_{k:T}\rangle\langle\psi_{k:T}|$ with $|\psi_{k:T}\rangle = U(x_T)\cdots U(x_k)|0\rangle$, and the *depth* of $\rho_{k:T}$ is $T - k + 1$ step unitaries.

Lemma 3.1. *Let $w_k = (1 - f_k)P_k$ for $1 \le k \le T$, with P_m as in (1.6). Then $w_k \ge 0$, and for every $0 \le m \le T$,*

$$\sum_{k=1}^{m} w_k = P_m. \quad (3.2)$$

In particular $\sum_{k=1}^{T} w_k = P_T = 1$, so $(w_k)_{k=1}^{T}$ is a probability vector.

Proof. Non-negativity is clear since $f_k \in [0,1]$. From (1.6), $P_{k-1} = f_k P_k$, so

$$w_k = (1 - f_k)P_k = P_k - f_k P_k = P_k - P_{k-1},$$

and (3.2) follows by telescoping, using $P_0 = 0$ from (1.4). Taking $m = T$ and $P_T = 1$ gives the last assertion. □

Proposition 3.2 (Renewal representation; cf. [16])**.** *Let ρ_T be given by Definition 1.1. Then*

$$\rho_T = \sum_{k=1}^{T} w_k\, \rho_{k:T}, \qquad w_k = (1 - f_k)P_k, \quad (3.3)$$

and consequently, with $g_k = \operatorname{Tr}(O\rho_{k:T})$,

$$y_T = \sum_{k=1}^{T} w_k\, g_k, \qquad |g_k| \le \|O\|. \quad (3.4)$$

Proof. The proof is by induction on T. For $T = 1$, Lemma 3.1 gives $w_1 = P_1 = 1$ and $\rho_{1:1} = \Phi_{x_1}(\sigma) = \rho_1$, which is (3.3).

Suppose (3.3) holds for $T - 1$, that is $\rho_{T-1} = \sum_{k=1}^{T-1} \widetilde{w}_k \rho_{k:T-1}$ with $\widetilde{w}_k = (1 - f_k)\widetilde{P}_k$ and $\widetilde{P}_m = \prod_{s=m+1}^{T-1} f_s$. By (1.5) and linearity of Φ_{x_T},

$$\begin{aligned} \rho_T &= \Phi_{x_T}\big(f_T \rho_{T-1} + (1 - f_T)\sigma\big) \\ &= f_T \sum_{k=1}^{T-1} \widetilde{w}_k\, \Phi_{x_T}(\rho_{k:T-1}) + (1 - f_T)\, \Phi_{x_T}(\sigma). \end{aligned}$$

By (3.1), $\Phi_{x_T}(\rho_{k:T-1}) = \rho_{k:T}$ for $k \le T - 1$, and $\Phi_{x_T}(\sigma) = \rho_{T:T}$. Moreover $f_T \widetilde{P}_k = P_k$ for $k \le T-1$, so $f_T \widetilde{w}_k = w_k$, while $(1 - f_T) = (1 - f_T)P_T = w_T$. Substituting gives (3.3) for T.

Equation (3.4) follows by linearity of the trace, and $|g_k| \le \|O\|\|\rho_{k:T}\|_1 = \|O\|$ by (2.1). □

Proposition 3.3 (Well-posedness)**.** *For every t, $\rho_t \in \mathcal{S}_d$, and $|y_t| \le \|O\|$.*

Proof. Φ_x is a unitary conjugation and $\mathcal{R}_f$ is a convex combination of the identity channel and the constant channel $\rho \mapsto \sigma$; both are completely positive and trace preserving, so $\mathcal{S}_d$ is invariant under (1.5). Alternatively, (3.3) exhibits ρ_T as a convex combination of states. The bound on y_t is (2.1). □

Proposition 3.3 is the structural difference from the classical cell: there the state c_t is an unconstrained vector in $\mathbb{R}^H$ whose norm may grow with t, whereas here the state space is compact and the output is bounded by $\|O\|$ irrespective of the inputs, the gate and the trainable unitaries.

4 Spectral structure

In this section the gate is taken to be constant, $f_t \equiv f \in [0,1)$ for $t \geq 2$, so that the weights w_k do not depend on the inputs. Define

$$\Omega = \Big\{ \textstyle\sum_{l=1}^{L} \sum_{q=1}^{n} s_{l,q}\lambda_q : \; s_{l,q} \in \{-1,0,1\} \Big\} \subset \mathbb{R}. \tag{4.1}$$

Theorem 4.1 (Frequency support and decay)**.** *Fix $1 \leq s \leq T$ and fix x_t for all $t \neq s$. Regard $h(\phi) = y_T\big|_{x_s=\phi}$ as a function of $\phi \in \mathbb{R}$. Then:*

(i) h is a generalised trigonometric polynomial, $h(\phi) = \sum_{\omega\in\Omega} c_\omega e^{i\omega\phi}$, with $c_{-\omega} = \overline{c_\omega}$.

(ii) If moreover $\lambda \in \mathbb{Z}^n$, so that $\Omega \subset \mathbb{Z}$ and h is 2π-periodic, then for every $\omega \in \Omega \setminus \{0\}$,

$$|c_\omega| \leq f^{\,T-s}\|O\|. \tag{4.2}$$

(iii) Jointly in all T inputs, y_T is a generalised trigonometric polynomial on $\mathbb{R}^T$ with frequency vectors in Ω^T.

Proof. (i) By (3.1) the variable x_s enters $\rho_{k:T}$ only through the factor $U(x_s)$, and only when $k \leq s$; for $k > s$ the state $\rho_{k:T}$ does not involve x_s at all. Write E_μ for the spectral projections of G, so that by (2.3) $S(\phi) = \sum_{\mu\in\mathrm{spec}(G)} e^{-i\mu\phi/2}E_\mu$. Substituting into (1.3), $U(\phi)$ is a finite sum of terms $e^{-i\phi(\mu_1+\cdots+\mu_L)/2}A$ with A independent of ϕ and $\mu_l \in \mathrm{spec}(G)$. In $g_k = \mathrm{Tr}(O\rho_{k:T})$ the unitary $U(\phi)$ occurs once and its adjoint once, so each resulting term carries a phase $e^{i\phi\theta}$ with

$$\theta = \tfrac{1}{2}\sum_{l=1}^{L}(\nu_l - \mu_l), \qquad \mu_l, \nu_l \in \mathrm{spec}(G).$$

By (2.3), $\frac{1}{2}(\nu_l - \mu_l) = \sum_q \lambda_q(s'_{l,q} - s_{l,q})/2$ with $s_{l,q}, s'_{l,q} \in \{-1,1\}$, and $(s'_{l,q} - s_{l,q})/2 \in \{-1,0,1\}$. Hence $\theta \in \Omega$. Summing over k with the constant weights w_k preserves this, and the conjugate symmetry holds because h is real valued.

(ii) By Proposition 3.2 and the observation in (i),

$$h(\phi) = \underbrace{\sum_{k\leq s} w_k\, g_k(\phi)}_{=:h_0(\phi)} + \underbrace{\sum_{k>s} w_k\, g_k}_{\text{independent of }\phi}.$$

The second sum contributes only to c_0, so for $\omega \neq 0$ the Fourier coefficients of h and of h_0 coincide. Since h_0 is 2π-periodic and continuous,

$$|c_\omega| = \left|\frac{1}{2\pi}\int_0^{2\pi} h_0(\phi)e^{-i\omega\phi}\,d\phi\right| \leq \sup_\phi |h_0(\phi)|$$
$$\leq \Big(\sum_{k\leq s} w_k\Big)\|O\| = P_s\|O\|,$$

using $|g_k| \leq \|O\|$ from (3.4) and (3.2). As the gate is constant, $P_s = f^{\,T-s}$, which is (4.2).

(iii) Apply the argument of (i) to each variable in turn; the phases multiply. □

Writing $m = T - s$ for the *age* of the input x_s, (4.2) says that the oscillatory dependence of the output on an input decays geometrically in the age of that input, at a rate set by the gate alone.

Remark 4.2. Part (iii) bounds the frequency *support*, not the number of free parameters: the coefficients c_ω are determined by $W_1, \ldots, W_L$, which in this experiment are parametrised by $2nL$ angles, and are therefore highly constrained. The numerical experiment of Section 7 shows that for $n = 2$, $L = 1$ the number of *active* frequencies grows as $2, 12, 88, 594$ for $T = 1, 2, 3, 4$ while the parameter count stays at 4; this is a statement about the support of the spectrum and should not be read as an assertion of 594 independent degrees of freedom.

Remark 4.3 (Classical containment)**.** If $\lambda \in \mathbb{Z}^n$ and the inputs are confined to a grid of P equally spaced points per period, the functions $\phi \mapsto e^{i\omega\phi}$ for ω and $\omega + P$ are indistinguishable, so the model is contained in the classical trigonometric model spanned by $\{e^{i\omega\phi} : |\omega| \leq \lfloor P/2 \rfloor\}$. This is the precise sense in which no representational advantage over a classical Fourier model is available on periodically sampled data.

5 Fading memory

Theorem 5.1 (Weight-independent fading Lipschitz bound)**.** *Let ρ_T, y_T be as in Definition 1.1 with $\varphi \in C^1$. Then for every $1 \leq s \leq T$,*

$$\left|\frac{\partial y_T}{\partial x_s}\right| \leq \|O\|\Big(\prod_{t=s+1}^{T} f_t\Big)\Big(\Lambda + 2\,|\varphi'(x_s)|\Big), \tag{5.1}$$

where by (1.4) the term $\varphi'(x_1)$ is to be read as 0. The right hand side does not depend on

$W_1, \dots, W_L$. *If* $L_\varphi = \sup_x |\varphi'(x)|$, *then*

$$\sum_{s=1}^{T} \left| \frac{\partial y_T}{\partial x_s} \right| \le \|O\| \big(\Lambda + 2L_\varphi\big) \sum_{s=1}^{T} P_s \le \frac{\|O\| \big(\Lambda + 2L_\varphi\big)}{1 - f_{\max}}, \tag{5.2}$$

uniformly in T.

Proof. Fix s and set $D_t = \partial \rho_t / \partial x_s$, a self-adjoint traceless element of $\mathcal{M}_d$ (traceless because $\operatorname{Tr} \rho_t \equiv 1$). For $t < s$, $D_t = 0$.

Step $t = s$. Write $M_s = \mathcal{R}_{f_s}(\rho_{s-1}) = f_s \rho_{s-1} + (1 - f_s)\sigma$, abbreviate $U = U(x_s)$, and note that ρ_{s-1} and σ do not depend on x_s while $f_s = \varphi(x_s)$ does. Differentiating $\rho_s = U M_s U^*$,

$$D_s = U' M_s U^* + U M_s (U')^* + \varphi'(x_s)\, U(\rho_{s-1} - \sigma) U^*,$$

where $U' = U'(x_s)$. By (2.1), Lemma 2.1, $\|M_s\|_1 = 1$ and $\|\rho_{s-1} - \sigma\|_1 \le 2$,

$$\begin{aligned} \|D_s\|_1 &\le 2\|U'\| \|M_s\|_1 + |\varphi'(x_s)| \|\rho_{s-1} - \sigma\|_1 \\ &\le \Lambda + 2|\varphi'(x_s)|. \end{aligned} \tag{5.3}$$

For $s = 1$ the gate term is absent by (1.4) and (5.3) holds with $\varphi'(x_1) = 0$.

Steps $t > s$. Here f_t, σ and $U(x_t)$ are independent of x_s, so differentiating (1.5) gives $D_t = f_t\, U(x_t) D_{t-1} U(x_t)^*$, whence by (2.1) and unitarity $\|D_t\|_1 = f_t \|D_{t-1}\|_1$. Iterating from $t = s+1$ to T,

$$\|D_T\|_1 = \Big(\prod_{t=s+1}^{T} f_t \Big) \|D_s\|_1 = P_s \, \|D_s\|_1.$$

Finally $|\partial y_T / \partial x_s| = |\operatorname{Tr}(O D_T)| \le \|O\| \|D_T\|_1$, which with (5.3) gives (5.1). The right hand side of (5.1) involves only $\|O\|$, the gate values and Λ, hence is independent of the W_l.

For (5.2), bound $|\varphi'(x_s)| \le L_\varphi$ and use $P_s \le f_{\max}^{T-s}$ to obtain $\sum_{s=1}^{T} P_s \le \sum_{j=0}^{T-1} f_{\max}^{j} < (1 - f_{\max})^{-1}$. □

Remark 5.2. No bound of the form (5.2) holds for the classical LSTM cell. For that cell the output is read off as $y = w_o^\top c_T$, so multiplying every trainable weight by α multiplies the read-out contribution to $\partial y / \partial x_s$ by α; no bound depending only on the architecture, the inputs and the gate can therefore hold. This is confirmed numerically in Section 7: the maximal total sensitivity of a classical cell grows from $1.2 \cdot 10^{-2}$ to $4.4 \cdot 10^{3}$ as α ranges over $[1/4, 16]$, whereas for the map (1.5) it stays below 5.5 against a bound of $\Lambda/(1-f) = 70$.

Remark 5.3. The sum $\sum_{s=1}^{T} P_s$ appearing in (5.2) is exactly the expected circuit depth of Theorem 6.1 below; the same quantity controls both the memory of the map and the cost of estimating its output.

The effect of discarding the remote past is quantified next. For $1 \le K \le T$ let $y_T^{(K)}$ denote the output of the map (1.5) run on the truncated input $(x_{T-K+1}, \dots, x_T)$, that is, re-initialised at σ at time $T - K + 1$.

Corollary 5.4 (Truncation)**.** *For every* $1 \le K \le T$,

$$\big| y_T - y_T^{(K)} \big| \le 2\|O\| \prod_{t=T-K+1}^{T} f_t \le 2\|O\| f_{\max}^{K}. \tag{5.4}$$

Proof. Apply Proposition 3.2 to both runs. The truncated run has weights w'_k for $T - K + 1 \le k \le T$; since the gate values f_t for $t \ge T - K + 2$ are unchanged, $w'_k = w_k$ for $k \ge T - K + 2$, while by Lemma 3.1 applied to the truncated sequence, $w'_{T-K+1} = P_{T-K+1}$. By (3.2), $\sum_{k \le T-K+1} w_k = P_{T-K+1}$ as well. Hence

$$\begin{aligned} y_T - y_T^{(K)} &= \sum_{k=1}^{T-K+1} w_k g_k - P_{T-K+1} g_{T-K+1} \\ &= \sum_{k=1}^{T-K+1} w_k \big(g_k - g_{T-K+1} \big). \end{aligned}$$

The $k = T - K + 1$ term vanishes, so the sum runs to $T - K$ and, using $|g_k - g_{T-K+1}| \le 2\|O\|$ and (3.2) once more,

$$\begin{aligned} \big| y_T - y_T^{(K)} \big| &\le 2\|O\| \sum_{k=1}^{T-K} w_k = 2\|O\| \, P_{T-K} \\ &= 2\|O\| \prod_{t=T-K+1}^{T} f_t. \end{aligned}$$

□

6 Unbiased estimation with bounded expected depth

Proposition 3.2 expresses y_T as an expectation over the index k, which suggests estimating it by sampling k and evaluating a single horizon state. The point of the following result is that the weights w_k are computable classically from the gate values alone, so the sampling may be done on a classical computer and each quantum circuit is a plain unitary circuit: no mid-circuit measurement, no conditional reset and no ancilla register is required.

Theorem 6.1 (Horizon sampling)**.** *Assume σ is pure and O is self-adjoint with $O^2 = I$, so that a measurement of O returns a value in $\{-1, +1\}$. Draw $k_1, \dots, k_N$ independently from the probability vector $(w_k)_{k=1}^T$ of Lemma 3.1, and for each j let $Z_j \in \{-1, +1\}$ be the outcome of a single measurement of O on the state $\rho_{k_j:T}$. Put $\widehat{y} = N^{-1} \sum_{j=1}^N Z_j$. Then:*

(i) $\mathbb{E}[\widehat{y}] = y_T$;

(ii) for every $\varepsilon > 0$, $\mathbb{P}\big(|\widehat{y} - y_T| \geq \varepsilon\big) \leq 2e^{-N\varepsilon^2/2}$;

(iii) the expected number of step unitaries in the sampled circuit is

$$\mathbb{E}[\,T - k + 1\,] = \sum_{m=1}^{T} P_m \;\leq\; \frac{1}{1 - f_{\max}}, \quad (6.1)$$

a bound independent of T.

Proof. (i) Conditionally on $k_j = k$, the measurement of O on $\rho_{k:T}$ has expectation $\mathrm{Tr}(O\rho_{k:T}) = g_k$. Hence $\mathbb{E}[Z_j] = \sum_k w_k g_k = y_T$ by (3.4), and averaging preserves this.

(ii) The Z_j are independent and take values in $[-1, 1]$, an interval of length 2. Hoeffding's inequality gives $\mathbb{P}(|\widehat{y} - \mathbb{E}\widehat{y}| \geq \varepsilon) \leq 2\exp(-2N\varepsilon^2/2^2) = 2e^{-N\varepsilon^2/2}$.

(iii) The depth of $\rho_{k:T}$ is $T - k + 1$, a positive integer, so by the tail formula for the expectation of a non-negative integer-valued random variable,

$$\begin{aligned}\mathbb{E}[T - k + 1] &= \sum_{m=1}^{T} \mathbb{P}(T - k + 1 \geq m) \\ &= \sum_{m=1}^{T} \mathbb{P}(k \leq T - m + 1) \\ &= \sum_{m=1}^{T} P_{T-m+1},\end{aligned}$$

using (3.2). Re-indexing by $j = T - m + 1$ gives $\sum_{j=1}^T P_j$, and $P_j \leq f_{\max}^{T-j}$ yields $\sum_{j=1}^T P_j \leq \sum_{i=0}^{T-1} f_{\max}^i < (1 - f_{\max})^{-1}$. □

Combining Theorem 6.1(iii) with Corollary 5.4 gives the practical statement: to estimate y_T to additive accuracy ε it suffices to run circuits of expected depth $(1 - f_{\max})^{-1}$, truncated at $K = \lceil \log(\varepsilon/2\|O\|)/\log f_{\max} \rceil$ steps, with $N = O(\varepsilon^{-2} \log \delta^{-1})$ shots — all three quantities independent of the sequence length T.

7 Numerical verification

All of the results of Sections 3–6 were verified numerically. Two independent implementations were used: a direct density-matrix simulator written from Definition 1.1, and the PennyLane library (version 0.45.1) [3], in which $\mathcal{R}_f$ is implemented as a channel with Kraus operators $K_0 = \sqrt{f}\, I$ and $K_{j+1} = \sqrt{1 - f}\, |0\rangle\langle j|$ for $0 \leq j \leq d - 1$, evaluated on the `default.mixed` device, while the horizon states (3.1) are evaluated on `default.qubit`. The two implementations agree to $9.4 \cdot 10^{-16}$ in trace distance.

Unless stated otherwise the configuration is $n = 3$, $L = 2$, $\lambda = (1, 2, 4)$ — so $\Omega = \{-14, \dots, 14\}$ and $\Lambda = 14$ — $T = 8$, $O = Z_1$ with $\|O\| = 1$, and $\varphi(x) = f_{\min} + (f_{\max} - f_{\min})\varsigma(ux + v)$ with ς the logistic function. Each bound was tested on 200 instances with $W_1, \dots, W_L$, the inputs, and the gate parameters drawn at random. Table 1 reports, for each result, the extreme value of the relevant quantity over all instances.

Two entries deserve comment. The ratio 0.257 for the decay estimate (4.2) and the ratio 0.347 for (5.1) indicate that the bounds are valid but not attained on random instances; whether they are sharp has not been investigated. The row marked

Table 1 Numerical verification. Rows marked "(PennyLane)" were computed in the independent implementation. Quantities in the third column are the worst case over 200 random instances; identities are reported as absolute errors and inequalities as the ratio of the measured quantity to its bound, which must not exceed 1. Here B_s denotes the right hand side of (5.1) and $m = T - s$ the age of the perturbed input.

Result	Quantity	Worst case	Threshold
Prop. 3.3	$\lvert \operatorname{Tr} \rho_t - 1 \rvert$	$2.44{\cdot}10^{-15}$	10^{-9}
	$\lambda_{\min}(\rho_t)$	$-3.71{\cdot}10^{-16}$	$\geq -10^{-9}$
	$\lvert \operatorname{Tr} \rho_T - 1 \rvert$ (PennyLane channel)	$4.23{\cdot}10^{-15}$	10^{-8}
Prop. 3.2	$\lvert y_T - \sum_k w_k g_k \rvert$	$5.55{\cdot}10^{-16}$	10^{-9}
	$\lvert \sum_k w_k - 1 \rvert$, $-\min_k w_k$	$2.22{\cdot}10^{-16}$	10^{-9}
	$\lvert y_T - \sum_k w_k g_k \rvert$ (PennyLane)	$8.05{\cdot}10^{-16}$	10^{-8}
Thm. 4.1	$\max_{\omega \notin \Omega} \lvert c_\omega \rvert$	$1.14{\cdot}10^{-16}$	10^{-12}
	$\max_{\omega \neq 0} \lvert c_\omega \rvert / f^m$	0.257	1
	$\max_{\omega \notin \Omega} \lvert c_\omega \rvert$ (PennyLane)	$7.11{\cdot}10^{-17}$	10^{-10}
Thm. 5.1	$\lvert \partial y_T / \partial x_s \rvert / B_s$	0.347	1
	$\lvert \partial y_T / \partial x_s \rvert / B_s$ (PennyLane)	0.187	1
	$\lvert \partial y_T / \partial x_s \rvert / B_s$, trained model	0.095	1
Cor. 5.4	$\lvert y_T - y_T^{(K)} \rvert \,/\, 2 \prod_{t>T-K} f_t$	0.691	1
Thm. 6.1	coverage of Hoeffding ball ($\delta = 0.05$)	0.990	≥ 0.95
	$\lvert$bias$\rvert$ in standard errors	2.170	< 4
	$\mathbb{E}[\text{depth}]\,(1 - f_{\max})$	0.869	1

"trained model" records that (5.1) continues to hold, with ratio 0.095, for a map whose parameters were fitted by gradient descent to real data rather than drawn at random, confirming that the estimate is not vacuous in the regime of interest.

For Remark 5.2, the trainable weights of a classical LSTM cell with $H = 4$ and those of the map (1.5) were scaled by a common factor α and the maximal total input sensitivity recorded over 40 random draws at each α (Figure 1).

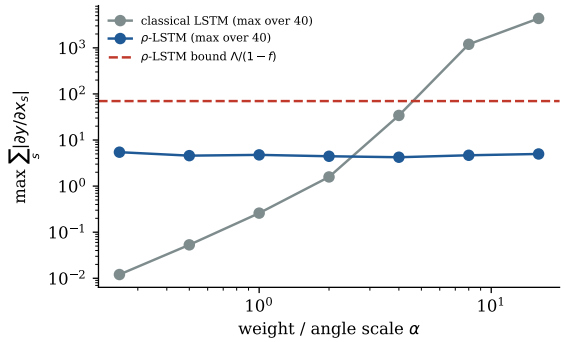


Fig. 1 Maximal total input sensitivity $\max \sum_s |\partial y / \partial x_s|$ against a common scaling α of the trainable weights, over 40 random draws at each α. The classical cell is unbounded in α; the map (1.5) remains below the bound $\Lambda/(1 - f)$ of (5.2), shown dashed

Finally, the map was fitted to a real engineering time series — the hourly oil temperature of an electricity transformer, from the publicly available ETT dataset of Zhou et al. [17] — predicting the next hourly increment from the previous 24, with a chronological split. With $n = 4$, $L = 2$ and 23 trainable parameters the test mean squared error was 0.2585 ± 0.0052 over three seeds, against 0.2312 ± 0.0287 for a classical LSTM with 35 parameters and 0.2098 for ridge regression on the same window; the learned gate corresponds to a memory length $(1 - \bar{f})^{-1}$ of 1.7 to 2.0 hours. These figures are recorded for completeness: the map is competitive but not superior, consistent with Remark 1.2, and the purpose of the experiment here is the verification of (5.1) on a trained model reported above.

8 Execution on quantum hardware

Proposition 3.2 replaces one channel circuit by T unitary circuits. Whether the resulting reconstruction survives device noise was tested. For a fixed input sequence with $T = 8$ and constant gate $f = 0.7$, the 8 horizon states (3.1) were prepared on two superconducting processors of IBM Quantum, `ibm_fez` and `ibm_kingston`, and

g_k estimated with 4096 shots per circuit using the Estimator primitive with readout-error mitigation (`resilience_level` $= 1$, twirled readout extinction). The circuits were checked against the reference simulator before submission, agreeing to $6.7 \cdot 10^{-16}$. Results are in Table 2.

In Table 2, $\overline{|\delta g|}$ is the mean absolute deviation of the measured g_k from their exact values, $\|w\|_2 \,\mathrm{sd}(\delta g)$ is the deviation predicted for the weighted sum if the per-term errors were independent, and $|\sum_k w_k \delta g_k|$ is the realised error of the reconstruction; the exact value is $y_T = 0.04498$. The individual horizon terms carry errors of 0.036 and 0.093 on the two processors, against a shot-noise standard deviation of 0.015, so device error dominates sampling error. The reconstructed output is nevertheless accurate to 0.012 and 0.034 respectively, and in both cases this agrees with the deviation predicted by treating the per-term errors as independent and weighting them by w: the convex weights of Proposition 3.2 average the per-circuit errors rather than accumulating them. The expected depth was 3.14 step unitaries against a maximum of 8, consistent with (6.1), whose bound here is $(1 - 0.7)^{-1} = 3.33$. A separate run of the sampling estimator of Theorem 6.1 with $N = 4000$ shots on `ibm_fez` returned $\widehat{y}$ within $1.1 \cdot 10^{-2}$ of y_T.

Remark 8.1. No measured resource advantage is claimed over a direct implementation of the channel (1.5), which would require mid-circuit measurement with conditional reset, or an ancilla register and a controlled swap; that comparison was not run. The comparison that the data support is the one in Table 2: the reconstruction is accurate, and the circuits required are shallow, unitary and reset-free.

9 Concluding remarks

The map of Definition 1.1 differs from a classical recurrent cell in three respects that are consequences of its state being a density matrix and its gate a channel: the state space is compact, so no normalisation is needed (Proposition 3.3); the recurrence decomposes exactly into finitely many unitary trajectories (Proposition 3.2, extending [16]); and the input sensitivity obeys a bound that does not involve the trainable parameters (Theorem 5.1). The third property is the one with no classical counterpart, and it is obtained at a cost: by Remark 4.3 the same structure that produces the bound also confines the map, on periodically sampled data, to a classical trigonometric model.

Several questions remain. Whether (4.2) or (5.1) is sharp is not known. The generalisation to non-pure reference states σ and to non-unitary Φ is straightforward for Proposition 3.2 but changes Theorem 6.1, since the horizon states are then mixed. Finally, it would be of interest to know for which classes of input sequences the containment of Remark 4.3 is strict.

Acknowledgements. The authors acknowledge the use of IBM Quantum services for the hardware results presented in this work. The views expressed are those of the authors and do not reflect the official policy or position of IBM or the IBM Quantum team.

Statements and Declarations

- **Funding.** This work was supported by the Maharashtra Quantum Mission, Government of Maharashtra, under the grant reference MQMDTEM-D170TEQP/60/2026-DESK17.
- **Competing interests.** The authors have no competing interests to declare that are relevant to the content of this article.
- **Ethics approval and consent to participate.** Not applicable.
- **Consent for publication.** Not applicable.
- **Data availability.** The numerical results, the time-series fitting results and the raw hardware estimates of the horizon terms are deposited in Zenodo at https://doi.org/10.5281/zenodo.23854141; the files are under embargo until publication of this article and are available to reviewers and readers on request.
- **Code availability.** The verification notebook, with the direct density-matrix simulator and the PennyLane implementation, is archived in the same Zenodo record and is available to reviewers and readers on request.
- **Author contribution.** All authors contributed to the study conception and design. P.P. developed the theory and proofs, wrote the verification code and carried out the hardware experiments. P.D. assisted with parts of the code. R.G. and R.A. carried out numerical checks and prepared the documentation. M.N.

Table 2 Reconstruction of y_T from hardware estimates of the horizon terms ($T = 8$, $f = 0.7$, 4096 shots per circuit).

Processor	$\overline{\lvert\delta g\rvert}$	$\lVert w\rVert_2\,\mathrm{sd}(\delta g)$	$\lvert\sum_k w_k \delta g_k\rvert$	$\mathbb{E}$[depth]
`ibm_fez`	0.0359	0.0091	0.0117	3.14
`ibm_kingston`	0.0933	0.0414	0.0343	3.14

and S.B. supervised the work. The first draft of the manuscript was written by P.P. and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

References

[1] Bausch, J.: Recurrent quantum neural networks. In: Advances in Neural Information Processing Systems, vol. 33, pp. 1368–1379 (2020). https://doi.org/10.48550/arXiv.2006.14619

[2] Berberich, J., Fink, D., Pranjić, D., Tutschku, C., Holm, C.: Training robust and generalizable quantum models. Phys. Rev. Res. **6**, 043326 (2024) https://doi.org/10.1103/PhysRevResearch.6.043326

[3] Bergholm, V., et al.: PennyLane: automatic differentiation of hybrid quantum-classical computations. arXiv:1811.04968 (2018). https://doi.org/10.48550/arXiv.1811.04968

[4] Carmichael, H.J.: An Open Systems Approach to Quantum Optics. Lecture Notes in Physics Monographs, vol. 18. Springer, Berlin (1993). https://doi.org/10.1007/978-3-540-47620-7

[5] Chen, J., Nurdin, H.I.: Learning nonlinear input-output maps with dissipative quantum systems. Quantum Inf. Process. **18**, 198 (2019) https://doi.org/10.1007/s11128-019-2311-9

[6] Chen, J., Nurdin, H.I., Yamamoto, N.: Temporal information processing on noisy quantum computers. Phys. Rev. Appl. **14**, 024065 (2020) https://doi.org/10.1103/PhysRevApplied.14.024065

[7] Chen, S.Y.-C., Yoo, S., Fang, Y.-L.L.: Quantum long short-term memory. In: ICASSP 2022 – IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 8622–8626 (2022). https://doi.org/10.1109/ICASSP43922.2022.9747369

[8] Čindrak, S., Donvil, B., Lüdge, K., Jaurigue, L.: Enhancing the performance of quantum reservoir computing and solving the time-complexity problem by artificial memory restriction. Phys. Rev. Res. **6**, 013051 (2024) https://doi.org/10.1103/PhysRevResearch.6.013051

[9] Dalibard, J., Castin, Y., Mølmer, K.: Wave-function approach to dissipative processes in quantum optics. Phys. Rev. Lett. **68**, 580–583 (1992) https://doi.org/10.1103/PhysRevLett.68.580

[10] Evans, M.R., Majumdar, S.N., Schehr, G.: Stochastic resetting and applications. J. Phys. A Math. Theor. **53**, 193001 (2020) https://doi.org/10.1088/1751-8121/ab7cfe

[11] Feller, W.: An Introduction to Probability Theory and Its Applications vol. I, 3rd edn. Wiley, New York (1968)

[12] Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997) https://doi.org/10.1162/neco.1997.9.8.1735

[13] Pérez-Salinas, A., Cervera-Lierta, A., Gil-Fuster, E., Latorre, J.I.: Data re-uploading for a universal quantum classifier. Quantum **4**, 226 (2020) https://doi.org/10.22331/q-2020-02-06-226

[14] Sannia, A., Martínez-Peña, R., Soriano, M.C., Giorgi, G.L., Zambrini, R.: Dissipation as a resource for quantum reservoir computing. Quantum **8**, 1291 (2024) https://doi.org/10.22331/q-2024-03-20-1291

[15] Schuld, M., Sweke, R., Meyer, J.J.: Effect of data encoding on the expressive power of

variational quantum machine learning models. Phys. Rev. A **103**, 032430 (2021) https://doi.org/10.1103/PhysRevA.103.032430

[16] Wald, S., Yao, L.H., Platini, T., Hooley, C., Carollo, F.: Stochastic resetting in discrete-time quantum dynamics: steady states and correlations in few-qubit systems. Quantum **9**, 1742 (2025) https://doi.org/10.22331/q-2025-05-13-1742

[17] Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., Zhang, W.: Informer: beyond efficient transformer for long sequence time-series forecasting. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, pp. 11106–11115 (2021). https://doi.org/10.1609/aaai.v35i12.17325