

Learning spectral density functions in open quantum systems

Felipe Peleteiro ¹, João Victor Shiguetsugo Kawanami Lima ¹, Pedro Marcelo Prado ¹,
Felipe Fernandes Fanchini ¹ and Ariel Norambuena ^{2,*}

¹*School of Sciences, São Paulo State University (UNESP), 17033-360 Bauru, São Paulo, Brazil*

²*Departamento de Física, Universidad Técnica Federico Santa María, Casilla 110 V, Valparaíso, Chile*



(Received 5 March 2026; revised 23 June 2026; accepted 17 July 2026; published 3 August 2026)

Spectral density functions quantify how environmental modes couple to quantum systems and govern their open dynamics. Inferring such frequency-dependent functions from time-domain measurements is an ill-conditioned inverse problem. Here, we use exactly solvable spin-boson models with pure-dephasing and amplitude-damping channels to reconstruct spectral density functions from noisy simulated data. First, we introduce a parameter estimation approach based on machine learning regressors to infer Lorentzian and Ohmic-like spectral density parameters, quantifying robustness to noise. Second, we show that a cosine transform inversion yields a physics-consistent spectral prior estimation, which is refined by a constrained neural network enforcing positivity and correct asymptotic behavior. Our neural-network framework robustly reconstructs structured spectral densities by filtering simulated noisy signals and learning general functional dependencies.

DOI: [10.1103/bqw4-317c](https://doi.org/10.1103/bqw4-317c)

I. INTRODUCTION

In the theory of open quantum systems, the microscopic description of the environment introduces bath frequencies ω_k and the corresponding system-bath coupling constants $g_k \in \mathbb{C}$, where k may denote a discrete or continuous index. The spectral density function (SDF) is defined as $J(\omega) = \sum_k |g_k|^2 \delta(\omega - \omega_k)$ [1,2], which encodes how environmental modes couple to the system under interest. The Dirac delta function $\delta(\omega - \omega_k)$ effectively models the environment as a continuous spectral distribution, facilitating analytical treatments for damping rates and time evolution. By construction, $J(\omega) \geq 0$, and the normalization condition $\int_0^\infty J(\omega) d\omega = \sum_k |g_k|^2$ must hold.

At low frequencies ($\omega \rightarrow 0$), density-of-states arguments or noise-correlation arguments typically lead to a power-law scaling $J(\omega) \propto \omega^s$ with $s \in \mathbb{R}^+$. Conversely, at high frequencies ($\omega \rightarrow \infty$), the SDF must decay, reflecting the suppression of high-energy bath excitations. The most widely used phenomenological model is the Ohmic-like SDF $J(\omega) = \alpha \omega_c^{1-s} \omega^s e^{-\omega/\omega_c}$, which captures the main physical features of a broad class of environments [3–5]. Here, $\alpha > 0$ sets the coupling strength, ω_c is the cutoff frequency, and s distinguishes sub-Ohmic ($s < 1$), Ohmic ($s = 1$), and super-Ohmic ($s > 1$) cases. While the Ohmic class enables analytical progress and provides valuable intuition, it fails to capture the complexity

of structured environments, such as quantum dots [6] or solid-state spin defects in diamond [7–9].

Moving beyond phenomenological models, one may use *ab initio* calculations [7,9] or molecular dynamics simulations [8,10]. However, these approaches are computationally demanding, susceptible to discretization artifacts, and sensitive to many-body correlations. Alternatively, microscopic arguments within exactly solvable models can be invoked, although such methods remain restricted to specific scenarios, for instance, light-matter interactions in the electric-dipole approximation [1].

A promising direction is to estimate $J(\omega)$ directly from experimental data by combining artificial intelligence (AI) with physically consistent quantum models. Given a set of measurements, one may infer the SDF that best explains the observed dynamics under suitable constraints. Current AI-assisted strategies primarily employ neural networks to classify [11–13] or fit parameters [12] within the Ohmic class for the pure-dephasing spin-boson model. However, a systematic AI framework capable of reconstructing structured environments beyond Ohmic-like parametrizations from noisy data remains lacking. Moreover, robustness analyses indicate that small deviations in the SDF can lead to exponentially amplified errors in expectation values, highlighting the intrinsic sensitivity of this inverse quantum problem [14].

In this work, we present two complementary approaches to learn SDFs from time-domain data in exactly solvable spin-boson models, which we use to benchmark our results. The first performs AI-based parameter estimation within phenomenological spectral density families using machine learning regressors. The second exploits a cosine transform structure to construct a nonparametric spectral prior, which is subsequently stabilized by a physics-constrained neural network. We benchmark accuracy and robustness across noise

*Contact author: ariel.norambuena@usm.cl

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

regimes and discuss how these tools extend to broader classes of open quantum systems with known models.

II. MODELS

First, we consider the pure-dephasing (PD) spin-boson model $\hat{H}_{\text{PD}} = \hat{H}_s + \hat{H}_b + \hat{V}_{\text{PD}}$, where $\hat{H}_s = (\omega_0/2)\hat{\sigma}_z \otimes \mathbb{1}_b$ describes a two-level system with Bohr frequency ω_0 and $\hat{H}_b = \mathbb{1}_s \otimes \sum_k \omega_k \hat{a}_k^\dagger \hat{a}_k$ is the bosonic bath. Here, $\hat{\sigma}_z = |1\rangle\langle 1| - |0\rangle\langle 0|$ is the Pauli matrix in the two-level-system basis $\{|0\rangle, |1\rangle\}$, and $\hat{a}_k$ ($\hat{a}_k^\dagger$) is the annihilation (creation) boson operator. The interaction Hamiltonian for the PD channel is given by ($\hbar = 1$)

$$\hat{V}_{\text{PD}} = \hat{\sigma}_z \otimes \sum_k (g_k \hat{a}_k + g_k^* \hat{a}_k^\dagger), \quad (1)$$

where g_k are the coupling constants and the SDF is defined as

$$J(\omega) = \sum_k |g_k|^2 \delta(\omega - \omega_k). \quad (2)$$

The PD model captures the orbital-phonon dynamics in color centers [15] and time-invariant quantum discord [16], both of which describe Markovian and non-Markovian regimes. For a thermal bath state $\hat{\rho}_b = e^{-\beta \hat{H}_b} / \text{Tr}[e^{-\beta \hat{H}_b}]$ with $\beta = (k_B T)^{-1}$ the inverse temperature, the exact time-local dynamics is $\dot{\hat{\rho}}_{\text{PD}} = (\gamma_{\text{PD}}(t)/2)[\hat{\sigma}_z \hat{\rho}_{\text{PD}} \hat{\sigma}_z - \hat{\rho}_{\text{PD}}]$ [16,17]. Consistently with the convention below, the dephasing rate is

$$\gamma_{\text{PD}}(t) = \dot{G}(t) = \int_0^\infty \frac{J(\omega)}{\omega} \coth\left(\frac{\beta\omega}{2}\right) \sin(\omega t) d\omega. \quad (3)$$

Here, $n(\omega) = [e^{\beta\hbar\omega} - 1]^{-1}$ is the mean boson number and $\coth(\beta\omega/2) = 2n(\omega) + 1$ is the total occupation factor for absorption and emission processes. Since populations remain constant in the PD model, the quantum coherence $C(t) = \sum_{i \neq j} |\rho_{ij}(t)|$ [18] is a natural observable. Experimentally, it can be reconstructed as

$$f_{\text{PD}}(t) = \sqrt{\langle \hat{\sigma}_x(t) \rangle^2 + \langle \hat{\sigma}_y(t) \rangle^2} = 2|\rho_{01}(0)|e^{-G(t)}, \quad (4)$$

$$G(t) = \int_0^\infty \frac{J(\omega)}{\omega^2} \coth\left(\frac{\beta\omega}{2}\right) [1 - \cos(\omega t)] d\omega, \quad (5)$$

where $f_{\text{PD}}(t) = C(t)$ is the quantum coherence signal for the PD model, with initial value $C(0) = 2|\rho_{01}(0)|$. Second, we consider the amplitude-damping (AD) spin-boson model, whose Hamiltonian is $\hat{H}_{\text{AD}} = \hat{H}_s + \hat{H}_b + \hat{V}_{\text{AD}}$, in which the system and bath Hamiltonians are the same as the PD model. In the rotating-wave approximation, the system-environment interaction Hamiltonian for the AD channel is given by ($\hbar = 1$)

$$\hat{V}_{\text{AD}} = \sum_k (g_k \hat{\sigma}_+ \otimes \hat{a}_k + g_k^* \hat{\sigma}_- \otimes \hat{a}_k^\dagger), \quad (6)$$

where $\hat{\sigma}_+ = (\hat{\sigma}_x + i\hat{\sigma}_y)/2$ and $\hat{\sigma}_- = (\hat{\sigma}_x - i\hat{\sigma}_y)/2$ describe excitation and relaxation of the two-level system, respectively.

We study spontaneous emission in the single-excitation manifold, where $\hat{N} = \hat{\sigma}_+ \hat{\sigma}_- + \sum_k \hat{a}_k^\dagger \hat{a}_k$ is the excitation number operator. Because $[\hat{N}, \hat{H}_{\text{AD}}] = 0$, the state for the initially excited system can be written as $|\Psi(t)\rangle = c_1(t)|1, \mathbf{0}\rangle +$

$\sum_k c_k(t)|0, 1_k\rangle$. The initial state is $|\Psi(0)\rangle = |1, \mathbf{0}\rangle$, and the reduced dynamics obeys

$$\begin{aligned} \dot{\hat{\rho}}_{\text{AD}} = & -iS(t)[\hat{\sigma}_+ \hat{\sigma}_-, \hat{\rho}_{\text{AD}}] \\ & + \gamma(t) \left[\hat{\sigma}_- \hat{\rho}_{\text{AD}} \hat{\sigma}_+ - \frac{1}{2} \{\hat{\sigma}_+ \hat{\sigma}_-, \hat{\rho}_{\text{AD}}\} \right], \end{aligned} \quad (7)$$

where $\hat{\rho}_{\text{AD}} = \text{Tr}_b[|\Psi(t)\rangle\langle\Psi(t)|]$, $S(t) = -\text{Im}[\dot{c}_1(t)/c_1(t)]$, and $\gamma(t) = -2\text{Re}[\dot{c}_1(t)/c_1(t)]$ [2]. Since populations evolve in this model, a suitable observable is

$$f_{\text{AD}}(t) = \langle \hat{\sigma}_z \rangle = 2|c_1(t)|^2 - 1, \quad (8)$$

$$\dot{c}_1(t) = - \int_0^t K(t-\tau) c_1(\tau) d\tau, \quad (9)$$

$$K(t) = \int_0^\infty e^{i(\omega_0 - \omega)t} J(\omega) d\omega. \quad (10)$$

Equations (4) and (8) define a map $J(\omega) \mapsto f_m(t)$ with a model label $m \in \{\text{PD}, \text{AD}\}$; throughout, we address the corresponding inverse problem $f_m(t) \mapsto J(\omega)$. In general, for different open quantum systems the role of the SDF into observables can be derived as explained in Appendix A. Maps in open quantum systems are useful, as we shall illustrate later, for training and guiding optimization in AI-based models, as shown in the schematic overview in Fig. 1.

III. PHENOMENOLOGICAL MODELS AND PARAMETER ESTIMATION

We introduce parametrized families of SDFs, denoted by $J(\omega, \xi)$, with $\xi \in \mathbb{R}^p$ representing model parameters, where $p \geq 1$ is the parameter dimension. In our computational simulations, we use two known SDFs,

$$J(\omega, \xi) = 2\alpha \omega_c^{1-s} \omega^s e^{-\omega/\omega_c} \quad \text{PD model}, \quad (11)$$

$$J(\omega, \xi) = \frac{1}{\pi} \frac{\gamma_0 \lambda^2}{(\omega - \omega_b)^2 + \lambda^2} \quad \text{AD model}, \quad (12)$$

where $\xi = [\alpha, s, \omega_c]$ corresponds to the Ohmic class and $\xi = [\gamma_0, \lambda, \omega_b]$ describes Lorentzian functions. Here, γ_0 is the amplitude, λ the width, and ω_b the bosonic frequency of the peak associated with the Lorentzian SDF. At zero temperature ($T = 0$), we obtain $f_{\text{PD}}(t) = 2|\rho_{10}(0)|\mathcal{F}_{\text{PD}}(t)$ and $f_{\text{AD}}(t) = 2\rho_{11}(0)\mathcal{F}_{\text{AD}}(t) - 1$, where analytical expressions for the signals are given by

$$\mathcal{F}_{\text{PD}}(t) = \exp \left[-2\alpha \Gamma(s) \int_0^{\omega_c t} \frac{\sin(s \tan^{-1} x)}{(1+x^2)^{s/2}} dx \right], \quad (13)$$

$$\mathcal{F}_{\text{AD}}(t) = e^{-\lambda t} \left| \cosh \left(\frac{dt}{2} \right) + \frac{\Lambda}{d} \sinh \left(\frac{dt}{2} \right) \right|^2, \quad (14)$$

where $d = [\Lambda^2 - 4\gamma_0\lambda]^{1/2}$, $\Lambda = \lambda - i\delta$, $\delta = \omega_0 - \omega_b$ is the detuning, and $\Gamma(s)$ is the Gamma function. As in the standard analytically solvable Lorentzian-reservoir model, the pole evaluation of the AD correlation function uses the continuum extension around the resonance. The analytical expression of $\mathcal{F}_{\text{PD}}(t)$ is derived in Ref. [16], while the expression for $\mathcal{F}_{\text{AD}}(t)$ includes the detuning δ . From Eqs. (13) and (14), variations in the model parameter vector ξ leave different, generally correlated signatures in the signal $f_m(t)$. Thus, estimating the

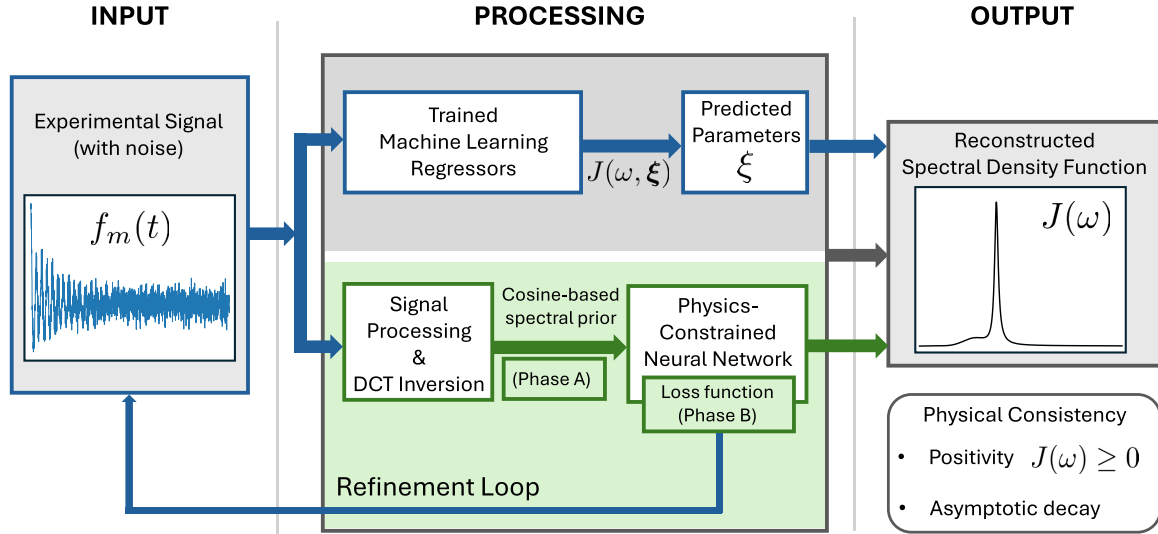


FIG. 1. Conceptual overview of the spectral density reconstruction problem. A quantum system coupled to a structured environment is characterized by a spectral density function $J(\omega)$, which determines measurable dynamical signals $f_m(t)$ through a known map $J(\omega) \mapsto f_m(t)$. The inverse problem, inferring $J(\omega)$ from noisy time-domain data, is ill-conditioned under finite time sampling and measurement noise. We address this challenge via two complementary AI-based strategies: (1) a parametric regression scheme for estimating SDF parameters and (2) a nonparametric approach combining discrete cosine transform (DCT) inversion with a constrained neural-network refinement to obtain physically consistent spectral reconstructions.

SDF from data depends on how sensitively the measured trace responds to ξ .

Previous work has quantified the effect of SDF variations on open-system observables [14]. Here, we specialize the discussion to the AD model. Let $J \mapsto J + \delta J$, let c_1 and $\tilde{c}_1$ denote the corresponding amplitudes, and set $a = \|J\|_1$ and $b = \|\delta J\|_1$, where $\|x\|_1 = \int_0^\infty |x(\omega)| d\omega$. Since $|K(t)| \leq a$, $|\delta K(t)| \leq b$, and physical survival amplitudes satisfy $|c_1(t)|, |\tilde{c}_1(t)| \leq 1$, the difference $u(t) = |\tilde{c}_1(t) - c_1(t)|$ obeys the comparison inequality

$$u(t) \leq \frac{bt^2}{2} + a \int_0^t (t - \tau) u(\tau) d\tau. \quad (15)$$

Solving the corresponding equality gives the rigorous upper bound

$$|\Delta f_{\text{AD}}(t)| \leq \frac{4b}{a} [\cosh(\sqrt{a}t) - 1], \quad (16)$$

with the continuous limit $|\Delta f_{\text{AD}}(t)| \leq 2bt^2$ when $a \rightarrow 0$.

For the positive-frequency Lorentzian used here, $a = (\gamma_0 \lambda / \pi) [\pi/2 + \arctan(\omega_b / \lambda)]$. Integrating Eq. (16) over $t \in [0, t_f]$ gives

$$\int_0^{t_f} |\Delta f_{\text{AD}}(t)| dt \leq \frac{4b}{a} \left[\frac{\sinh(\sqrt{a}t_f)}{\sqrt{a}} - t_f \right]. \quad (17)$$

Equation (17) provides a finite-time worst-case stability bound. It guarantees continuity of the forward map in the L^1 norm, while its growth with the observation window also illustrates why inversion becomes progressively sensitive to noise and parameter degeneracy.

Simulated data and machine learning regressors

Simulated data for both PD and AD models are generated by sampling physical regimes in which the signals (13)

and (14) are well defined. To mimic relative measurement fluctuations, we use bounded multiplicative Gaussian noise, $f_m(t_i) \rightarrow f_m(t_i)[1 + \delta_i]$, with $\delta_i \sim \mathcal{N}(0, \sigma^2)$ independently at each time and clipped to $[-2\sigma, 2\sigma]$. The observable is subsequently clipped to its physical range. Each simulated trace is represented by the vector $\mathbf{x}^{(r)} = [f_m(t_1), \dots, f_m(t_{N_t})]^T$ and paired with its generating parameters $\xi^{(r)}$. The supervised dataset is therefore $\mathcal{D}_m = \{(\mathbf{x}^{(r)}, \xi^{(r)})\}_{r=1}^{N_s}$. For a regressor R_θ that maps a complete time trace to the SDF parameters, training minimizes

$$\mathcal{L}_{\text{par}}(\theta) = \frac{1}{N_s} \sum_{r=1}^{N_s} \|R_\theta(\mathbf{x}^{(r)}) - \xi^{(r)}\|^2. \quad (18)$$

In the implementation, one regressor is fitted for each component of ξ . The analytical maps (13) and (14) are used to generate labeled traces, while the trained regressors solve the inverse map directly. A pedagogical description and model hyperparameters are given in Appendix B.

For the AD channel, we benchmark support vector regression (SVR), a multilayer perceptron (MLP), random forest (RF), extreme gradient boosting (XGBoost), and histogram-based gradient boosting regression (HistGBR). We generated 6000 signals on a common grid of 200 times in $t \in [0, 10]$, using $\omega_0 = 1$ and independently sampling $(\lambda, \gamma_0, \omega_b)$ uniformly from $[0.1, 1]$. For each of three random train-test splits, 75% of the traces were used for training and 25% for testing. Models trained on clean signals were evaluated at 0%, 1%, 5%, and 10% test noise, using five independent noise realizations at each nonzero level. In a second protocol, noise-aware training was implemented by assigning each training trace an independent noise standard deviation sampled uniformly from $[0, 0.1]$.

Figure 2 shows the mean-square error (MSE) of the different clean-trained models to estimate SDF parameters as a function of the noise amplitude. We observe that RF yields the

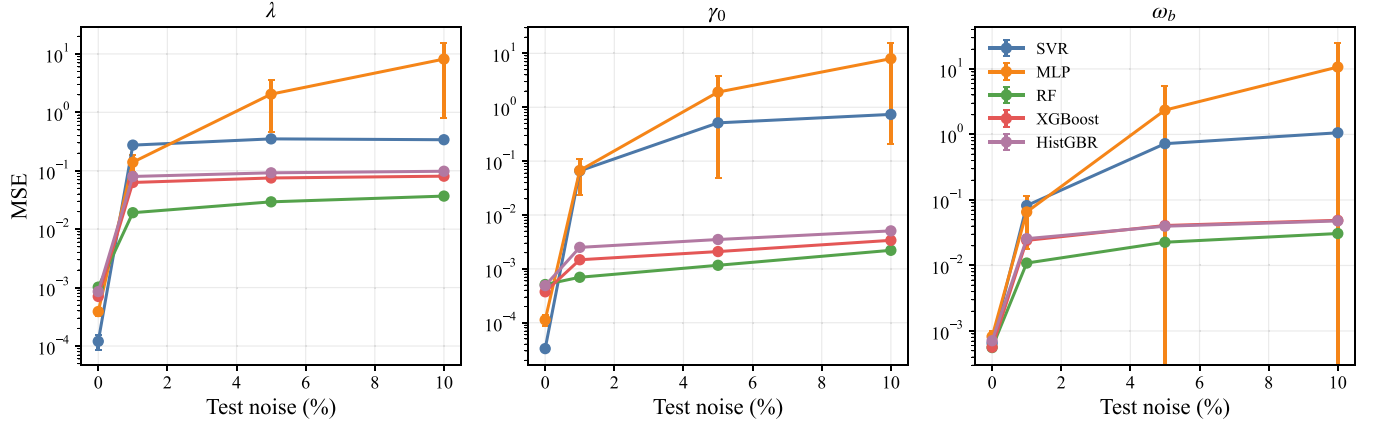


FIG. 2. Noise robustness of parametric SDF inference in the amplitude-damping (AD) channel. From left to right, the panels show the MSE for λ , γ_0 , and ω_b for models trained on clean signals and tested at 0%, 1%, 5%, and 10% multiplicative noise. Markers and error bars show means and standard deviations over three train-test splits and, at each nonzero noise level, five independent noise realizations. The vertical axes are logarithmic.

smallest errors in all cases, and between parameters, (λ , γ_0) are easier to detect than ω_b . This behavior is consistent with the structure of the forward map: λ and γ_0 jointly control the decay envelope and memory time, whereas ω_b is encoded through detuning-dependent oscillations that are easily obscured by noise. A finite-difference Jacobian analysis over 1000 random parameter sets gives median normalized sensitivities 0.379, 0.415, and 0.184 for λ , γ_0 , and ω_b , respectively. The median Jacobian condition number is 9.17, with a maximum 1.64×10^3 , confirming substantial local degeneracies.

Noise-aware training effectively mitigates this failure mode, as demonstrated in Fig. 3. At 5% test noise, the SVR MSE decreases from 0.351 to 1.56×10^{-3} for λ , from 0.514 to 2.11×10^{-4} for γ_0 , and from 0.723 to 1.88×10^{-3} for ω_b . The corresponding MLP errors decrease from 2.05, 1.91, and 2.35 to 3.25×10^{-3} , 8.39×10^{-4} , and 3.22×10^{-3} , respectively. Tree ensemble methods also show improvement for the more weakly identifiable λ and ω_b parameters, while RF performance changes minimally for γ_0 due to its inherent robustness. Therefore, the observed degradation is not an

intrinsic property of the regressors but rather results from a mismatch between clean training data and noisy observations.

IV. NEURAL-NETWORK LEARNING WITH PRIOR COSINE TRANSFORM

When the SDF parametrization is unknown, neural networks (NNs) can be used as universal function approximators [19,20]. Using a simulated dataset $\mathcal{D}_m$ and the known quantum map, we can train a deep learning model to infer SDFs from dynamical signals, even in the presence of noise. We now focus on the PD model and the structured SDF

$$J(\omega) = J_{\text{bulk}}(\omega) + J_{\text{loc1}}(\omega) + J_{\text{loc2}}(\omega) \quad [15], \text{ with}$$

$$J_{\text{bulk}}(\omega) = 2\alpha \omega_c^{1-s} \omega^s e^{-\omega/\omega_c}, \quad (19)$$

$$J_{\text{loc1}}(\omega) = \frac{J_0 \omega^s}{(\omega/\omega_{\text{loc}} + 1)^2} \frac{\Gamma/2}{(\omega - \omega_{\text{loc}})^2 + (\Gamma/2)^2}, \quad (20)$$

$$J_{\text{loc2}}(\omega) = J_1 \omega^s e^{-(\omega - \omega_0)^2/(2\sigma^2)}, \quad (21)$$

where $s = 3$, $\omega_c = 1$ THz, $\alpha = 0.0275$, $J_1 = 0.0025$ THz $^{-2}$, $\sigma = 2.4042$ THz, $\omega_0 = 9.35$ THz, $J_0 = 0.0235$ THz $^{-1}$, $\Gamma = 0.8414$ THz, and $\omega_{\text{loc}} = 15.19$ THz. The structured SDF accurately captures the electron-phonon coupling of the negatively charged silicon-vacancy center in diamond [8]. However, because of the complexity of the SDF (19)–(21), no closed analytical expression is available for the coherence signal $f_{\text{PD}}(t)$; only an approximate expression can be obtained by exploiting the pronounced peak of $J_{\text{loc1}}(\omega)$ [8]. The main idea is therefore to find a linear integral transformation between the function $F(f_{\text{PD}}(t))$ and the SDF $J(\omega)$, valid for any structured environment. Here, $F(\cdot)$ is a functional operator acting on the signal $f_{\text{PD}}(t)$.

In the PD model, the signal $H(t) = -d^2[\ln f_{\text{PD}}(t)]/dt^2$, with $f_{\text{PD}}(t) = C(0) \exp[-G(t)]$, satisfies the linear integral relation $H(t) = \int_0^\infty J(\omega) \coth(\beta\omega/2) \cos(\omega t) d\omega$, which has the structure of a cosine transform. Here, the operator is $F(\cdot) = -d^2[\ln(\cdot)]/dt^2$. This motivates introducing the half-range cosine transform $\mathcal{C}\{f\}(\Omega) = \int_0^\infty f(t) \cos(\Omega t) dt$, which is linear because $\mathcal{C}\{\alpha_1 f_1 + \alpha_2 f_2\}(\Omega) = \alpha_1 \mathcal{C}\{f_1\}(\Omega) +$

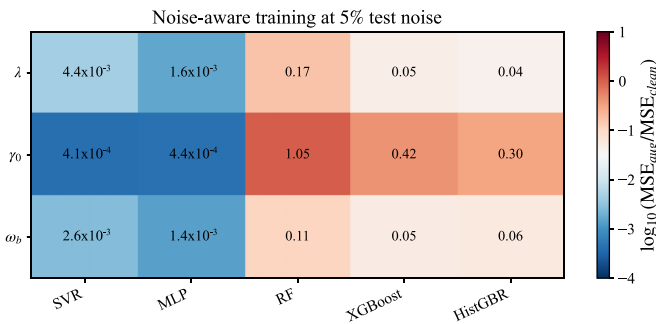


FIG. 3. Effect of noise-aware training at 5% test noise. Rows correspond to λ , γ_0 , and ω_b , and columns correspond to the five regression models. Cell annotations give $\text{MSE}_{\text{augmented}}/\text{MSE}_{\text{clean}}$, while the color scale represents the base-10 logarithm of this ratio. Values below one indicate improvement from noise-aware training. Results are averaged over three train-test splits and five independent test-noise realizations.

$\alpha_2 \mathcal{C}\{f_2\}(\Omega)$ for $\alpha_i \in \mathbb{C}$ and functions $f_i(t)$ whose cosine transforms are well defined. The inverse transform is $\mathcal{C}^{-1}\{F\}(t) = (2/\pi) \int_0^\infty F(\Omega) \cos(\Omega t) d\Omega$, which directly yields

$$J(\Omega) = \frac{2}{\pi} \tanh\left(\frac{\beta\Omega}{2}\right) \mathcal{C}\left\{\frac{d^2 G}{dt^2}\right\}(\Omega), \quad (22)$$

subject to the physical constraint $J(\Omega) \geq 0$. To derive Eq. (22), we have used the orthogonality relation $\int_0^\infty \cos(\omega t) \cos(\Omega t) dt = \pi \delta(\Omega - \omega)/2$, where $\delta(x)$ is the Dirac delta.

Equation (22) reveals a theoretical procedure to reconstruct the SDF from $G''(t) = -d^2[\ln f_{\text{PD}}(t)]/dt^2$ and the thermal factor $\tanh(\beta\Omega/2)$. Experimental data, however, consist of discrete noisy time samples. We therefore define an equally spaced time grid $t_n = n\Delta t$ ($n = 0, \dots, N-1$) with $t_f = (N-1)\Delta t$, for which the quadrature $H_{t_f}(v_k) = \Delta t \sum_{n=0}^{N-1} H(t_n) \cos(v_k t_n)$ with $v_k = \pi k/t_f$ provides a discrete cosine transform estimate. We obtain

$$H_{t_f}(v) = \int_0^\infty J(\omega) \coth\left(\frac{\beta\omega}{2}\right) K_{t_f}(v, \omega) d\omega, \quad (23)$$

$$K_{t_f}(v, \omega) = \frac{\sin[(v - \omega)t_f]}{2(v - \omega)} + \frac{\sin[(v + \omega)t_f]}{2(v + \omega)}. \quad (24)$$

In the asymptotic limit $t_f \rightarrow \infty$ and in the absence of aliasing ($\omega_{\text{max}} < \pi/\Delta t$), this kernel reduces to $H_{t_f}(v) \simeq (\pi/2)J(v) \coth(\beta v/2)$, so the spectral peaks of $H_{t_f}(v)$ and $J(v)$ coincide up to thermal weighting. These observations motivate a frequency-domain preprocessing stage before applying a deep learning algorithm, but are physically inspired by the model. Naturally, for a different open quantum system, a linear integral relation $H(t) \mapsto J(\omega)$ must be derived, taking into account specific details of the model. We then define the dataset $\mathcal{D}_\omega \equiv \{\mathcal{S}(\omega_n), \omega_n\}$ with $\mathcal{S}(\omega_n) = \{2/[\pi \coth(\beta\omega_n/2)]\} H_T(\omega_n)$. This provides a coarse reconstruction of $J(\omega)$ and a starting point for NN optimization. The NN is first pretrained by minimizing the loss function (phase A)

$$\mathcal{L}_{\text{pre}} = \frac{1}{L} \sum_{n=1}^L [J_{\text{NN}}(\omega_n; \theta) - \mathcal{S}(\omega_n)]^2, \quad (25)$$

yielding optimal NN parameters θ^* after training and fine tuning. In a second postprocessing stage, the reconstruction of the SDF is refined using time domain information encoded in the signal $f_m(t)$, by minimizing the loss function (phase B)

$$\mathcal{L}_{\text{post}} = \frac{1}{N} \sum_{i=1}^N [\hat{f}_m(t_i, J_{\text{NN}}(\omega; \varphi)) - [f_m(t_i)(1 + \delta_i)]]^2, \quad (26)$$

initialized at $\varphi = \theta^*$ obtained in the pretraining phase. Here, $\hat{f}_m[t_i, J_{\text{NN}}(\omega; \varphi)]$ is computed from the known quantum map $J(\omega) \mapsto f_m(t)$. Physical consistency is enforced by parametrizing

$$J_{\text{NN}}(\omega; \varphi) = g \mathcal{F}(\omega) \text{softplus}[\mathbf{N}_J(\omega; \varphi)], \quad (27)$$

where $g > 0$ is a learned gain, $\mathcal{F}(\omega) = \omega^d e^{-\omega/\Omega}$ is normalized by its maximum on the numerical grid, and $\text{softplus}(x) = \ln(1 + e^x)$ is the softplus function used in machine learning. This parametrization guarantees non-negativity and imposes

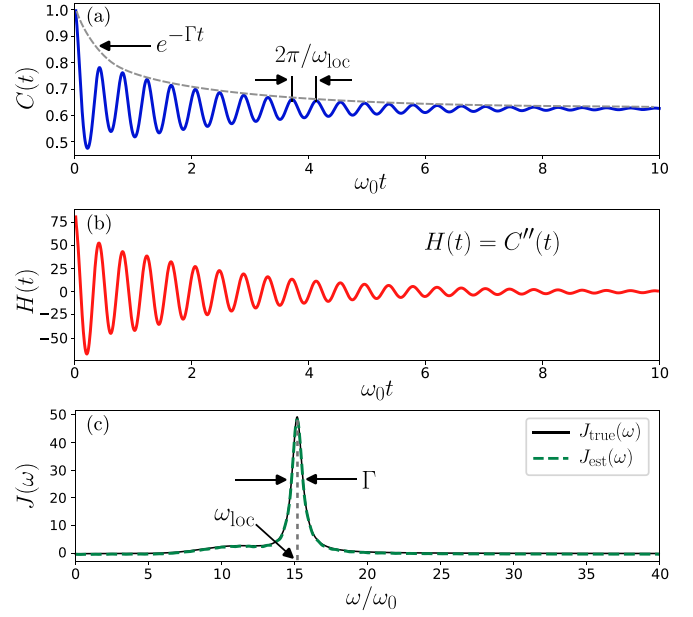


FIG. 4. Noise-free reconstruction using the discrete cosine transform. (a) Coherence signal generated from the phenomenological SDF in Eqs. (19)–(21). (b) Signal $H(t) = G''(t)$ used in the cosine inversion of Eq. (22). (c) True SDF and its discrete-cosine-transform estimate, without neural-network refinement.

the required low- and high-frequency behavior. More details of the NN protocol are given in Appendix C.

V. RESULTS OF SDF RECONSTRUCTION FOR THE PD MODEL USING NEURAL-NETWORK REFINEMENT

In the ideal noise-free setting (Fig. 4), the cosine transform in Eq. (22) already provides an accurate and parameter-free SDF reconstruction: it reproduces both the super-Ohmic envelope and the dominant localized features of $J(\omega)$ at negligible computational cost. This establishes a useful baseline: whenever the model is well known and the signal-to-noise ratio is high, a linear inversion can provide a good approximation to the SDF without imposing a phenomenological SDF. The present formulation therefore broadens previous approaches [11–13], which focused on noise-free scenarios and phenomenological SDF families.

The situation changes in the presence of noise. We perturb the coherence as $C(t) \rightarrow C(t)[1 + \delta(t)]$, producing the signal observed in Fig. 5(a). The discrete cosine transform then yields a nonphysical SDF with fictitious oscillations and negative regions, as shown in Fig. 5(c). This is expected because the second derivative in $H(t) = -d^2[\ln f_{\text{PD}}(t)]/dt^2$ amplifies short-time fluctuations and finite-window artifacts. We mitigate this instability by combining the cosine-based spectral prior (phase A) with a physics-constrained NN refinement (phase B). To quantify the SDF reconstruction, we use the error metrics

$$\varepsilon_J = \frac{\|J_{\text{rec}} - J_{\text{true}}\|_2}{\|J_{\text{true}}\|_2}, \quad \varepsilon_C = \frac{\|C_{\text{rec}} - C_{\text{true}}\|_2}{\|C_{\text{true}}\|_2}, \quad (28)$$

together with the dominant-peak position and full-width-at-half-maximum errors and the fraction of negative spectral

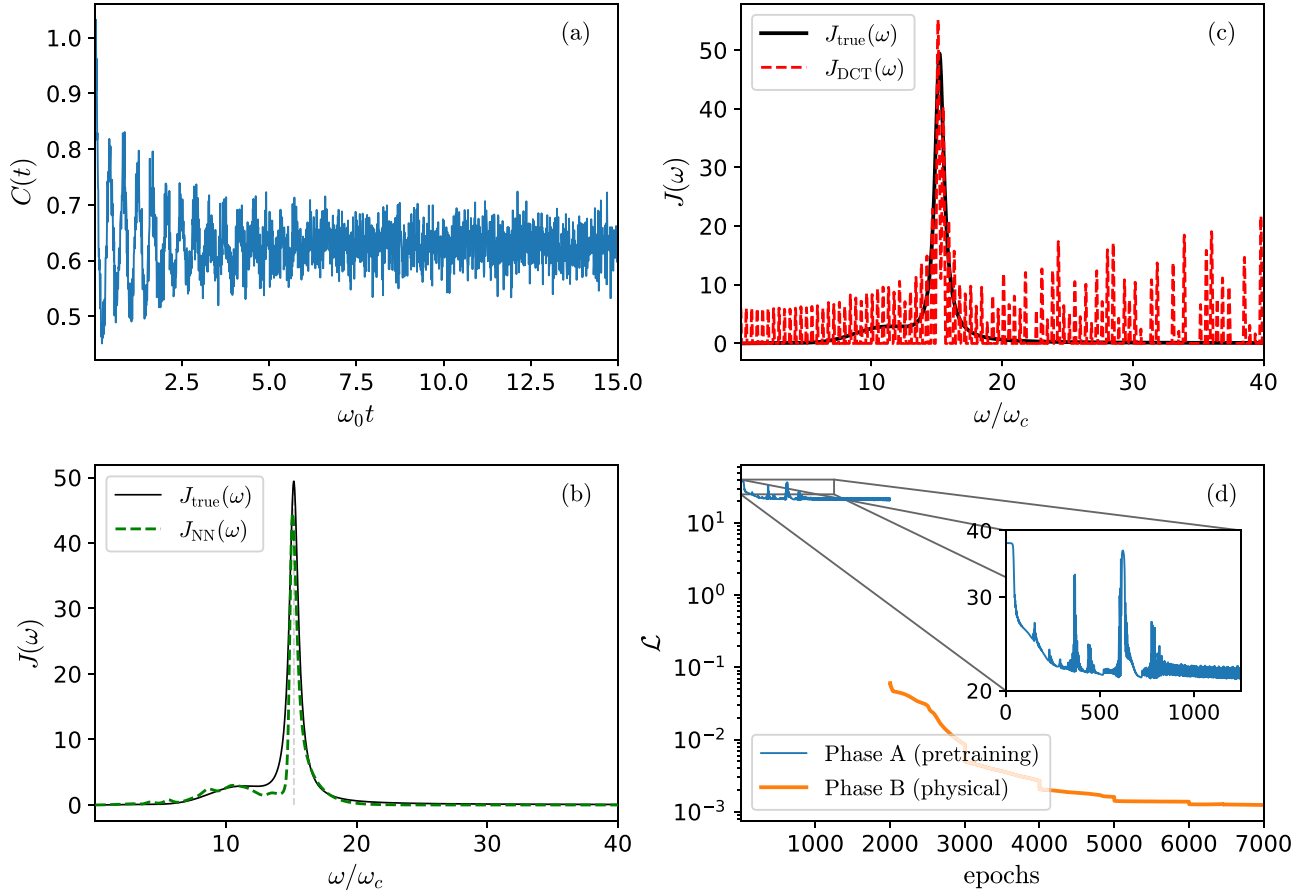


FIG. 5. Reconstruction of the SDF using a coherence signal with 5% noise. (a) Noisy coherence generated from the phenomenological model. (b) Neural-network reconstruction of the SDF using $J_{DCT}(\omega)$ as the preprocessing prior. (c) True SDF and its DCT estimate. (d) Loss convergence during preprocessing (phase A) and postprocessing (phase B), in which the SDF is refined by the neural network.

weight. Figure 6 reports means and standard deviations over five independent noise and training seeds. We compare the raw

DCT, a non-negative second-order Tikhonov inversion with its regularization strength selected on held-out time points [21],

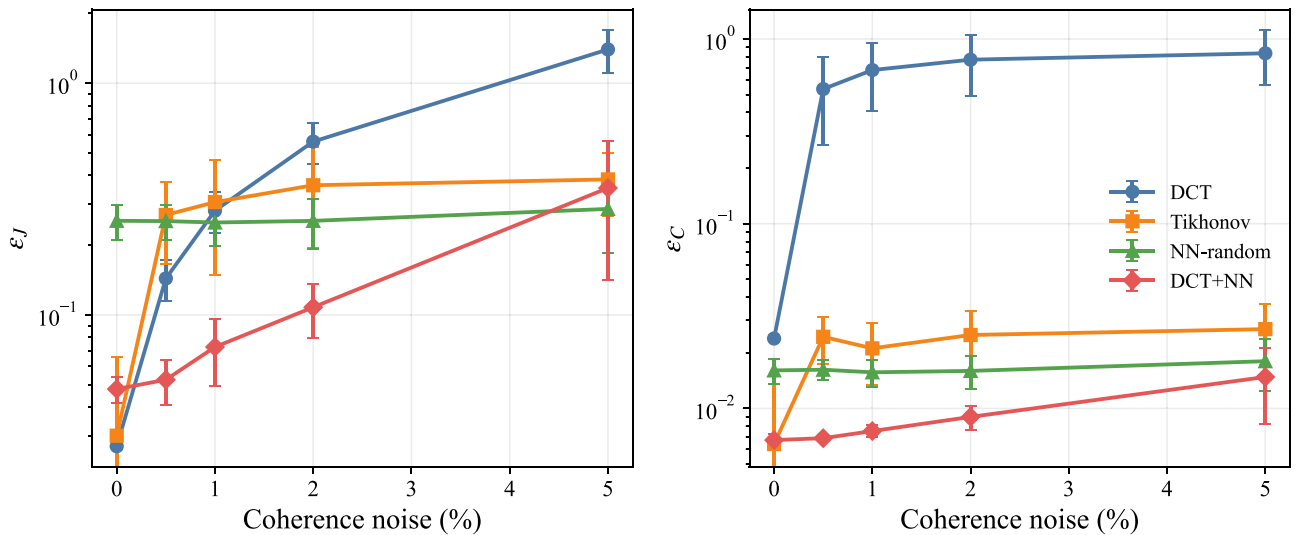


FIG. 6. Quantitative ablation of the pure-dephasing reconstruction. (Left) Relative spectral error ϵ_J and (right) coherence error ϵ_C vs multiplicative coherence noise for the raw DCT, non-negative Tikhonov inversion, randomly initialized constrained NN, and DCT-pretrained constrained NN. Markers and error bars show means and standard deviations over five independent noise and training seeds. Both vertical axes are logarithmic.

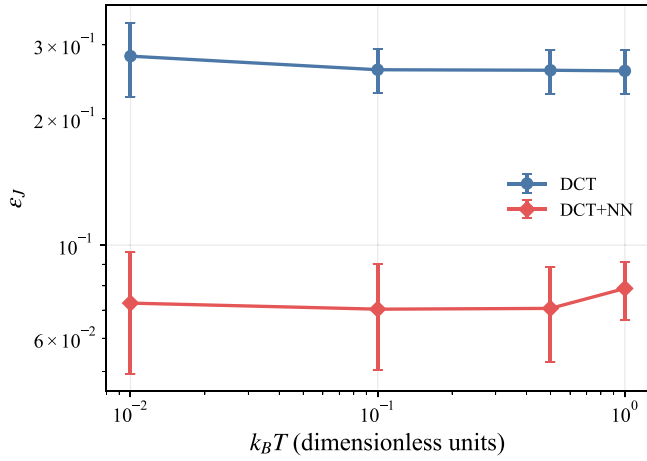


FIG. 7. Relative spectral error ε_J vs the known dimensionless temperature $k_B T$ at 1% coherence noise for the raw DCT and DCT-pretrained neural-network reconstructions. Markers and error bars show means and standard deviations over three independent seeds. Both axes are logarithmic.

the constrained NN trained from random initialization, and the full DCT + NN protocol presented in this work.

In the noise-free limit, the DCT gives $\varepsilon_J = 0.027$, and additional NN processing is unnecessary. At 0.5%, 1%, and 2% noise, the raw-DCT errors increase to 0.144 ± 0.029 , 0.282 ± 0.056 , and 0.560 ± 0.112 , whereas the full DCT + NN errors remain 0.053 ± 0.012 , 0.073 ± 0.023 , and 0.108 ± 0.028 . The NN trained from random initialization remains near $\varepsilon_J \simeq 0.25$, demonstrating the value of spectral pretraining in this moderate-noise regime. At 5% noise, the DCT fails ($\varepsilon_J = 1.40 \pm 0.29$ and a negative-weight fraction 0.52 ± 0.19). Both NN variants regularize the inversion; their global spectral errors overlap within the run-to-run variation, but the DCT-pretrained NN has a smaller coherence error (0.0148 versus 0.0180) and a much smaller relative peak-width error (0.080 versus 0.648). The classical Tikhonov baseline is competitive, especially in the noise-free limit, but at 1%–2% noise, its mean spectral error is approximately three to four times larger than that of DCT + NN. The NN should therefore be viewed as an adaptive physics-constrained regularizer, not as a universal replacement for classical inverse methods.

Temperature enters through $\coth(\beta\omega/2)$ in the forward map and $\tanh(\beta\omega/2)$ in the inversion. Figure 7 shows that for known temperature and the structured spectrum considered here, the DCT + NN spectral error remains between 0.070 and 0.079 over $k_B T \in [0.01, 1]$, while the DCT error remains near 0.26–0.28. If β is not independently calibrated, however, it becomes partially degenerate with the low-frequency behavior of $J(\omega)$ and should be inferred jointly using additional observables or informative priors.

These results show that the proposed framework regularizes the ill-conditioned inversion within the tested model class and controlled noise range; it does not remove the intrinsic ill-conditioning. The preprocessing remains channel specific. For PD, the logarithm and second derivative expose a linear cosine relation. For AD, $J(\omega)$ enters a memory kernel and then a nonlinear Volterra equation, so there is no equally

direct DCT prior. Parametric regression can nevertheless be used for PD when a reliable SDF family is known, and a nonparametric NN can be used for AD by differentiating through a Volterra solver. Broadband backgrounds may also be combined with parametric components while localized residual peaks are represented by a constrained NN or sparse basis. Experimental applications will additionally require calibrated colored-noise, state-preparation and measurement, and model-mismatch terms in the likelihood.

VI. CONCLUSIONS

We developed and benchmarked two complementary, data-driven routes to reconstruct spectral density functions in exactly solvable open quantum models. Inverting time-domain signals to obtain frequency-dependent SDFs is an ill-conditioned problem. Small perturbations in the data can lead to large deviations in the reconstructed spectra. This behavior is expected for compact integral operators (see Ref. [14] for related bounds).

In the parametric regime, regression techniques provide fast, interpretable estimates of SDF parameters, but their apparent noise robustness depends strongly on whether the training distribution includes measurement fluctuations. Noise augmentation reduces the largest AD regression errors by orders of magnitude. This route nevertheless assumes prior knowledge of the phenomenological family. In the nonparametric regime, the cosine-transform structure of the quantum map induces a physics-consistent spectral prior when the signal-to-noise ratio is sufficiently high. A constrained neural-network refinement then acts as a structured regularization mechanism, with a clear advantage over raw DCT and random initialization at low-to-moderate noise and a more limited, feature-dependent advantage at 5% noise.

From a mathematical perspective, the quantum map between dynamical observables and the SDF involves integral relations that define compact operators in frequency space. Their inversion amplifies high-frequency components and is intrinsically ill-conditioned under finite-time sampling and noisy data. The cosine prior, Tikhonov baseline, and constrained NN should therefore be understood as alternative regularization mechanisms with different bias-variance trade-offs. The present results provide a reproducible benchmark for environmental spectroscopy of structured reservoirs; validation with real experimental noise and model mismatch remains an essential next step.

ACKNOWLEDGMENTS

F.P., J.V.S.K.L., P.M.P., and F.F.F. acknowledge support from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Projects No. 2024/20249-8, No. 2024/15741-0, No. 2025/08578-9, and No. 2023/04987-6. A.N. acknowledges support from Fondecyt Regular No. 1251131, Fondecyt Exploración No. 13250014, and Anillo Temático ATE 250066.

DATA AVAILABILITY

The data and code supporting the findings of this article are publicly available in Ref. [22].

APPENDIX A: GKSL DERIVATION AND DEPENDENCE ON THE SDF

The starting point of the theory of open quantum systems is the Hamiltonian of the full (closed) system

$$\hat{H} = \hat{H}_s \otimes \mathbb{1}_b + \mathbb{1}_s \otimes \hat{H}_b + \hat{V}, \quad (\text{A1})$$

with $\hat{H}_s = \sum_a \hbar \omega_a |a\rangle\langle a|$, $\hat{H}_b = \sum_k \hbar \omega_k \hat{a}_k^\dagger \hat{a}_k$, and $\hat{V} = \hat{S} \hat{X}^\dagger + \hat{S}^\dagger \hat{X}$.

Here, we are only considering a collective bath operator of the form $\hat{X} = \sum_k g_k \hat{a}_k$, where g_k encodes the system-bath couplings. For initially uncorrelated states $\hat{\rho}(0) = \hat{\rho}_s(0) \otimes \hat{\sigma}_b$, we assume a bath density matrix in thermal equilibrium

$$\hat{\sigma}_b = \frac{e^{-\beta \hat{H}_b}}{\text{Tr}[e^{-\beta \hat{H}_b}]}, \quad \beta = (k_B T)^{-1}, \quad (\text{A2})$$

which maximizes the von Neumann entropy for a fixed average energy. Under the weak-coupling and secular approximations, the reduced state $\hat{\rho}_s(t) = \text{Tr}_b[\hat{\rho}(t)]$ follows the Gorini-Kossakowski-Sudarshan-Lindblad (GKSL) equation ($\hbar = 1$),

$$\begin{aligned} \dot{\hat{\rho}}_s = & -i[\hat{H}_s + \hat{H}_{\text{LS}}(t), \hat{\rho}_s(t)] \\ & + \sum_{\Omega, n} \gamma_n(t) \left[\hat{S}_n(\Omega) \hat{\rho}_s(t) \hat{S}_n^\dagger(\Omega) \right. \\ & \left. - \frac{1}{2} \{ \hat{S}_n^\dagger(\Omega) \hat{S}_n(\Omega), \hat{\rho}_s(t) \} \right]. \end{aligned} \quad (\text{A3})$$

Here, $\hat{H}_{\text{LS}}(t)$ denotes the Lamb-shift correction, and the system transition operators are

$$\hat{S}_n(\Omega) = \sum_{\omega_b - \omega_a = \Omega} |a\rangle\langle a| \hat{S}_n |b\rangle\langle b|. \quad (\text{A4})$$

For a positive Bohr frequency Ω , the second-order time-local absorption and emission rates have the transparent form

$$\gamma_\uparrow(\Omega, t) = 2 \int_0^\infty J(\omega) n(\omega) \frac{\sin[(\omega - \Omega)t]}{\omega - \Omega} d\omega, \quad (\text{A5})$$

$$\gamma_\downarrow(\Omega, t) = 2 \int_0^\infty J(\omega) [n(\omega) + 1] \frac{\sin[(\omega - \Omega)t]}{\omega - \Omega} d\omega, \quad (\text{A6})$$

where the ratio is understood by continuity at $\omega = \Omega$. Thus, the reduced density matrix naturally depends on $J(\omega)$, motivating the inverse problem of reconstructing the SDF from observables $\langle \hat{O} \rangle = \text{Tr}_s[\hat{\rho}_s(t) \hat{O}]$.

APPENDIX B: MACHINE LEARNING MODELS FOR PARAMETER ESTIMATION

1. Learning task in plain language

For the parametric method, a single input example is an entire population trace sampled at the same $N_t = 200$ times,

$$\mathbf{x} = [f_{\text{AD}}(t_1), \dots, f_{\text{AD}}(t_{N_t})]^\top. \quad (\text{B1})$$

The desired output is one physical parameter of the Lorentzian reservoir, namely λ , γ_0 , or ω_b . We train three separate regressors for each model family, one per target. This is analogous to learning three calibrated measuring instruments: each instrument reads the same time trace but is optimized to report a different physical quantity. Input standardization is used for SVR and the MLP, whereas tree models operate directly on the sampled signal values.

2. Why several regressors are compared

a. Support vector regression

SVR predicts a parameter from a weighted combination of representative training traces [23]. Its radial-basis kernel can describe smooth nonlinear relations while the ϵ -insensitive loss ignores sufficiently small residuals. This often yields excellent accuracy for clean, low-dimensional inverse maps, but a perturbation distributed across all time samples can change many kernel distances simultaneously.

b. Multilayer perceptron

The MLP is a feed-forward neural network that repeatedly applies affine maps and nonlinear activation functions. It learns a global smooth approximation to the inverse map and can combine information from the decay envelope and oscillatory features. Its flexibility is useful, but without representative noise during training it may interpret measurement fluctuations as physical structure.

c. Random forest

RF averages many decision trees trained on bootstrap samples and randomly selected subsets of input features [24]. Each tree partitions the space of time traces into regions with similar target values. Averaging reduces variance and makes the model comparatively insensitive to isolated distortions of individual time points.

d. Gradient-boosted trees

XGBoost and HistGBR build trees sequentially, with each new tree correcting residual errors left by the previous ensemble [25,26]. XGBoost uses explicit row and feature subsampling together with regularization. HistGBR first bins continuous inputs into histograms, which makes training efficient and introduces an additional coarse graining that can be helpful in noisy settings.

3. Exact implementation and hyperparameters

The benchmark uses $N_s = 6000$ traces with $(\lambda, \gamma_0, \omega_b)$ independently sampled from $[0.1, 1]$, a 75/25 train-test split, three independent split seeds, and five independent noise realizations for every nonzero test-noise level. Noise-aware training assigns each training trace a standard deviation drawn uniformly from $[0, 0.1]$. The exact estimators are:

(1) SVR: radial basis kernel, $C = 30$, $\epsilon = 0.01$, and $\text{gamma} = \text{scale}$, preceded by feature standardization.

(2) MLP: hidden widths (192,96), ReLU activations, Adam optimizer, L^2 penalty 10^{-4} , initial learning rate 8×10^{-4} , at most 450 iterations, a 15% internal validation subset,

and early stopping after 30 nonimproving iterations; inputs are standardized.

(3) RF: 350 trees, 70% of features considered at each split, one minimum sample per leaf, bootstrap aggregation, and all CPU cores.

(4) XGBoost: 550 trees of maximum depth 6, learning rate 0.045, row subsampling 0.85, column subsampling 0.8, L^2 regularization 1, squared-error objective, and histogram tree construction.

(5) HistGBR: at most 350 boosting iterations, learning rate 0.07, at most 31 leaves per tree, L^2 regularization 10^{-3} , and automatic early stopping.

Performance is reported using

$$\text{MSE}_q = \frac{1}{N_{\text{test}}} \sum_r (\hat{\xi}_q^{(r)} - \xi_q^{(r)})^2, \quad (\text{B2})$$

$$\text{MAE}_q = \frac{1}{N_{\text{test}}} \sum_r |\hat{\xi}_q^{(r)} - \xi_q^{(r)}| \quad (\text{B3})$$

for $q \in \{\lambda, \gamma_0, \omega_b\}$. The random seeds, fitted configurations, raw predictions summarized by run, and plotting routines are provided in Ref. [22].

APPENDIX C: ALGORITHMIC PROTOCOL FOR RECONSTRUCTING SPECTRAL DENSITY FUNCTIONS USING COSINE TRANSFORM

The reported PD benchmark uses $N = 800$ equally spaced times on $t \in [0, 15]$ and $M = 400$ frequencies on $\omega \in [10^{-3}, 40]$. More precisely, the full coherence signal is generated on the $N = 800$ -point time grid, while a uniformly subsampled set of 500 time points is used in the optimization losses to reduce computational cost. We set $\rho_{eg}(0) = 0.5$, use single-precision tensors, and evaluate the frequency integrals with trapezoidal weights. Independent random seeds 0, ..., 4 generate the noise and initialize the networks. The temperature sweep uses seeds 0, 1, 2. The neural-network stage should be understood as a physics-constrained reconstruction scheme: physical information is imposed through the positivity-preserving output, the asymptotic filter, and the forward open-system map, rather than through a differential-equation residual as in conventional physics-informed neural networks.

1. Signal and cosine prior

The noisy coherence is clipped to $[10^{-8}, 1]$ before computing $G(t) = -\ln \{C(t)/[2\rho_{eg}(0)]\}$. First and second derivatives are evaluated with second-order finite differences, and Eq. (22) is integrated by the trapezoidal rule. The raw DCT is retained for the reported baseline and negativity metric. For NN pretraining only, negative values are set to zero, positive outliers are capped at the 99.5th percentile, and an 11-point third-order Savitzky-Golay filter is applied. Thus, the DCT output plays two distinct roles: the unmodified DCT is used as a diagnostic baseline, whereas a clipped and smoothed version is used only as a stable spectral prior for phase A.

2. Network architecture

$\mathbf{N}_J(\omega)$ is a fully connected feed-forward network with one scalar input, four hidden layers of 128 neurons, hyperbolic-tangent activations, and one scalar output. Including the learned positive gain, the model has 49 922 trainable parameters. We use $\mathcal{F}(\omega) = \omega^3 \exp(-\omega/\Omega)$ with $\Omega = 4$, normalized by its maximum on the frequency grid, and a softplus output to enforce positivity. The factor $\mathcal{F}(\omega)$ fixes the expected low-frequency scaling and suppresses unphysical high-frequency spectral weight, while the neural network learns the remaining structured frequency dependence.

More explicitly, with normalized input $z_0 = 2\omega/\omega_{\text{max}} - 1$, the network evaluates

$$\mathbf{z}_\ell = \tanh(\mathbf{W}_\ell \mathbf{z}_{\ell-1} + \mathbf{b}_\ell), \quad \ell = 1, \dots, 4, \quad (\text{C1})$$

$$r(\omega) = \mathbf{W}_5 \mathbf{z}_4 + \mathbf{b}_5, \quad (\text{C2})$$

$$\mathcal{F}(\omega) = \frac{\omega^3 e^{-\omega/4}}{\max_{\omega_n} (\omega_n^3 e^{-\omega_n/4})}, \quad (\text{C3})$$

$$J_{\text{NN}}(\omega) = \text{softplus}(g_{\text{raw}}) \mathcal{F}(\omega) \text{softplus}[r(\omega)]. \quad (\text{C4})$$

The layer dimensions are $1 \rightarrow 128 \rightarrow 128 \rightarrow 128 \rightarrow 128 \rightarrow 1$. All weights and biases, together with the scalar g_{raw} , are trainable. PyTorch default linear-layer initialization is used after fixing the stated random seed.

3. Optimization

Let $\Delta^2 J_n = J_{n+1} - 2J_n + J_{n-1}$ and $S_J = \max(10^{-6}, \max_n J_n)$. The scale-invariant curvature penalty is

$$\mathcal{R}_{\text{curv}} = \frac{1}{M-2} \sum_{n=2}^{M-1} \left(\frac{\Delta^2 J_n}{S_J} \right)^2. \quad (\text{C5})$$

Phase A uses AdamW with learning rate 2×10^{-3} , weight decay 10^{-6} , and at most 350 epochs to minimize

$$\mathcal{L}_A = \text{SmoothL1} \left(\frac{J_{\text{NN}}}{S_{\text{prior}}}, \frac{J_{\text{prior}}}{S_{\text{prior}}} \right) + 2 \times 10^{-5} \mathcal{R}_{\text{curv}}, \quad (\text{C6})$$

where $S_{\text{prior}} = \max(1, \max_n J_{\text{prior},n})$. Here, SmoothL1 denotes the mean smooth- L_1 loss over the frequency grid, using the default PyTorch transition parameter. Phase B uses AdamW with learning rate 7×10^{-4} and at most 700 epochs to minimize

$$\begin{aligned} \mathcal{L}_B = & \frac{1}{N_{\text{tr}}} \sum_i \left[\frac{C_{\text{NN}}(t_i) - C_{\text{obs}}(t_i)}{S_C} \right]^2 \\ & + 2 \times 10^{-5} \mathcal{R}_{\text{curv}} \\ & + \frac{10^{-2} d_e}{M S_A^2} \sum_n (J_{\text{NN},n} - J_{A,n})^2, \end{aligned} \quad (\text{C7})$$

where $S_C = \max[10^{-4}, \sqrt{\langle C_{\text{obs}}^2 \rangle}]$, $S_A = \max(1, \max_n J_{A,n})$, and $d_e = \max[0.1, 1 - e/(699)]$ decays with epoch e . The quantity J_A denotes the spectral reconstruction obtained at the end of phase A; the last term in $\mathcal{L}_B$ therefore acts as a weak, epoch-decaying prior that stabilizes the early stage of the time-domain refinement without forcing the final solution to remain fixed at the DCT-pretrained spectrum. The learning

rate is halved after 80 epochs without improvement, with floor 2×10^{-5} . Early stopping is triggered after 100 and 160 non-improving epochs in phases A and B, respectively. Gradients are clipped at Euclidean norm 5.

4. Baselines and diagnostics

The random-initialization ablation uses the same architecture, physical filter, phase-B loss, and stopping criteria but omits phase A and the prior term. The Tikhonov baseline minimizes the time-domain residual with non-negativity

and a second-difference penalty; its strength is selected from $\{10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1\}$ using a held-out 20% subset of the time samples. All methods are evaluated using ε_J , ε_C , spectral R^2 , dominant-peak position and width, negative spectral weight, and runtime. Complete scripts, seeds, raw per-run metrics, aggregate tables, figure data, and environment specifications are available in Ref. [22]. These diagnostics are intended to quantify both spectral accuracy and dynamical consistency: ε_J and the peak metrics assess the reconstructed SDF itself, whereas ε_C measures whether the learned spectrum reproduces the observed open-system coherence.

-
- [1] I. de Vega and D. Alonso, Dynamics of non-Markovian open quantum systems, *Rev. Mod. Phys.* **89**, 015001 (2017).
 - [2] H.-P. Breuer and F. Petruccione, *The Theory of Open Quantum Systems* (Oxford University, Oxford, 2002).
 - [3] A. O. Caldeira and A. J. Leggett, Path integral approach to quantum Brownian motion, *Physica A* **121**, 587 (1983).
 - [4] A. J. Leggett, S. Chakravarty, A. T. Dorsey, M. P. A. Fisher, A. Garg, and W. Zwerger, Dynamics of the dissipative two-state system, *Rev. Mod. Phys.* **59**, 1 (1987).
 - [5] U. Weiss, *Quantum Dissipative Systems*, 3rd ed. (World Scientific, Singapore, 2008).
 - [6] I. Wilson-Rae and A. Imamoglu, Quantum dot cavity-QED in the presence of strong electron-phonon interactions, *Phys. Rev. B* **65**, 235311 (2002).
 - [7] A. Alkauskas, B. B. Buckley, D. D. Awschalom, and C. G. Van de Walle, First-principles theory of the luminescence lineshape for the triplet transition in diamond NV centres, *New J. Phys.* **16**, 073026 (2014).
 - [8] A. Norambuena, S. A. Reyes, J. Mejía-López, A. Gali, and J. R. Maze, Microscopic modeling of the effect of phonons on the optical properties of solid-state emitters, *Phys. Rev. B* **94**, 134305 (2016).
 - [9] M. C. Cambria, A. Norambuena, H. T. Dinani, G. Thiering, A. Gardill, I. Kemeny, Y. Li, V. Lordi, A. Gali, J. R. Maze, and S. Kolkowitz, Temperature-dependent spin-lattice relaxation of the nitrogen-vacancy spin triplet in diamond, *Phys. Rev. Lett.* **130**, 256903 (2023).
 - [10] F. J. González, D. Tancara, H. T. Dinani, R. Coto, and A. Norambuena, Engineering non-Markovianity from defect-phonon interactions, *New J. Phys.* **25**, 043004 (2023).
 - [11] J. Barr, G. Zicari, A. Ferraro, and M. Paternostro, Spectral density classification for environment spectroscopy, *Mach. Learn.: Sci. Technol.* **5**, 015043 (2024).
 - [12] J. Barr, S. Mukherjee, A. Ferraro, M. Paternostro, and G. Zicari, A machine learning based approach to the identification of spectral densities in quantum open systems, *Eur. Phys. J. Spec. Top.* **235**, 2977 (2026).
 - [13] J. Barr, A. Ferraro, M. Paternostro, and G. Zicari, Machine learning-enhanced characterisation of structured spectral densities: Leveraging the reaction coordinate mapping, [arXiv:2501.07485](https://arxiv.org/abs/2501.07485).
 - [14] F. Mascherpa, A. Smirne, S. F. Huelga, and M. B. Plenio, Open systems with error bounds: Spin-boson model with spectral density variations, *Phys. Rev. Lett.* **118**, 100401 (2017).
 - [15] A. Norambuena, J. R. Maze, P. Rabl, and R. Coto, Quantifying phonon-induced non-Markovianity in color centers in diamond, *Phys. Rev. A* **101**, 022110 (2020).
 - [16] P. Haikka, T. H. Johnson, and S. Maniscalco, Non-Markovianity of local dephasing channels and time-invariant discord, *Phys. Rev. A* **87**, 010103(R) (2013).
 - [17] J. Łuczka, Spin in contact with thermostat: Exact reduced dynamics, *Physica A* **167**, 919 (1990).
 - [18] T. Baumgratz, M. Cramer, and M. B. Plenio, Quantifying coherence, *Phys. Rev. Lett.* **113**, 140401 (2014).
 - [19] G. Cybenko, Approximation by superpositions of a sigmoidal function, *Math. Control Signals Syst.* **2**, 303 (1989).
 - [20] K. Hornik, Approximation capabilities of multilayer feedforward networks, *Neural Networks* **4**, 251 (1991).
 - [21] P. C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion* (SIAM, Philadelphia, 1998).
 - [22] F. Peleteiro, J. V. S. K. Lima, P. M. Prado, F. F. Fanchini, and A. Norambuena, Learning spectral density functions in open quantum systems: Code and reproducibility data, GitHub repository, 2026, <https://github.com/Ariel-Norambuena/learning-spectral-density-functions>.
 - [23] C. Cortes and V. Vapnik, Support-vector networks, *Mach. Learn.* **20**, 273 (1995).
 - [24] L. Breiman, Random forests, *Mach. Learn.* **45**, 5 (2001).
 - [25] J. H. Friedman, Greedy function approximation: A gradient boosting machine, *Ann. Stat.* **29**, 1189 (2001).
 - [26] T. Chen and C. Guestrin, XGBoost: A scalable tree boosting system, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (ACM, New York, 2016), pp. 785–794.