Universidade Federal de São Carlos
Centro de Ciências Exatas e de Tecnologia
Departamento de Física

Rodrigo Pereira Silva

EMPLOYING PHYSICS-INFORMED NEURAL NETWORKS FOR FINDING EFFECTIVE HAMILTONIANS IN QUANTUM SYSTEMS

São Carlos – SP
2026

Rodrigo Pereira Silva

EMPLOYING PHYSICS-INFORMED NEURAL NETWORKS FOR FINDING EFFECTIVE HAMILTONIANS IN QUANTUM SYSTEMS

Dissertação apresentada ao Programa de Pós-Graduação em Física da Universidade Federal de São Carlos, para obtenção do título de mestre em física.

Orientador: Prof. Dr. Celso Jorge Villas-Bôas

São Carlos – SP
2026

Silva, Rodrigo Pereira

Employing physics-informed neural networks for finding effective Hamiltonians in quantum systems / Rodrigo Pereira Silva -- 2026.
80f.

Dissertação (Mestrado) - Universidade Federal de São Carlos, campus São Carlos, São Carlos
Orientador (a): Celso Jorge Villas-Bôas
Banca Examinadora: Rodrigo Alves Dias, Victor Lopez Richard
Bibliografia

1. Mecânica quântica. 2. Aprendizado de máquina. 3. Redes neurais informadas pela física. I. Silva, Rodrigo Pereira. II. Título.

Ficha catalográfica desenvolvida pela Secretaria Geral de Informática (SIn)

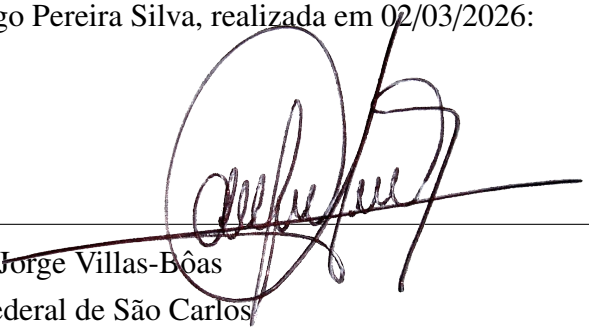
DADOS FORNECIDOS PELO AUTOR

Bibliotecário responsável: Arildo Martins - CRB/8 7180

UNIVERSIDADE FEDERAL DE SÃO CARLOS
Centro de Ciências Exatas e de Tecnologia
Programa de Pós-Graduação em Física

Folha de aprovação

Assinatura dos membros da comissão examinadora que avaliou e aprovou a Defesa de Mestrado do candidato Rodrigo Pereira Silva, realizada em 02/03/2026:



Prof. Dr. Celso Jorge Villas-Bôas
Universidade Federal de São Carlos

Documento assinado digitalmente
gov.br
VICTOR LOPEZ RICHARD
Data: 14/03/2026 18:34:25-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. Victor Lopez Richard
Universidade Federal de São Carlos

Documento assinado digitalmente
gov.br
RODRIGO ALVES DIAS
Data: 13/03/2026 09:12:18-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. Rodrigo Alves Dias
Universidade Federal de Juiz de Fora

Dedico este trabalho a todos que passaram por minha vida e me guiaram até aqui; e também a todos aqueles que, nas páginas que seguem, encontrarem utilidade para sua própria jornada.

Agradecimentos

Carrego comigo a sorte em ter o apoio de tantas pessoas, mas em momentos como este em que preciso agradecer, me vejo no dilema de me estender demais, ou de ser injusto com alguém. Depois de muito ponderar sobre o assunto, decidi por ser injusto com todos igualmente, e falar muito menos sobre cada um do que vocês realmente merecem. Aqui vão meus agradecimentos.

Agradeço primeiramente aos meus pais, Luiz e Luciene, que desde cedo investiram em minha formação e que sempre abriram mão de tudo que lhes era conforto para oferecer o melhor a mim. O sacrifício deles me rendeu a maior de todas as heranças: o conhecimento. Dedico este trabalho primordialmente aos dois. Eu não seria absolutamente nada sem vocês.

Agradeço à minha família: meus irmãos Renan, Eduardo e Fernando, que me apoiam mesmo que nem sempre me compreendam. Também agradeço à minha avó, dona Alfina, que é o maior exemplo de resiliência que já tive o privilégio de conhecer; a senhora é uma inspiração para mim. Agradeço também a todos os meus tios e primos que não vou citar um a um, pois corro o risco de precisar de um documento somente com seus nomes.

Agradeço aos meus amigos do peito, que não são muitos, e por isso sou feliz. Ricardo, Mateus, Cairo, Leo, dedico este parágrafo aos senhores. Aprendi com vocês o que significa dar asas à curiosidade e perseguir sonhos. Agradeço também aos meus amigos fora de casa, pois vocês se tornaram minha família por longos anos, depois que decidi ser cientista. Vou deixar o receio de lado e citar nomes, na esperança de não esquecer ninguém: Rodrigo, Felipe, Gabriel, Melissa, Marina, Lara, Mariza, Renne, Pierre, e todos vocês do ReSolvay, em especial ao Alberto, cuja amizade vivo ostentando por aí.

Agradeço também ao meu orientador, Celso, que foi uma das pessoas mais humanas que já tive o prazer de conhecer na área acadêmica. Também agradeço ao Gubio, que me guiou por esse caminho difícil de pesquisa, e que me ajudou como se o projeto fosse dele próprio. Sem vocês essas páginas não teriam sido escritas.

Agradeço à Catharina, que esteve comigo durante boa parte dessa aventura, e que foi minha companheira durante momentos em que sinceramente pensei em desistir. Durante esses momentos de dúvida, sempre pude recorrer à sua paciência e compreensão, na certeza de que seria ouvido de verdade. Você me mostrou o que significa parceria. Te amo.

E também não deixo de lado as pessoas que fizeram parte da minha outra jornada: Clarissa, Ciro, Pedro, Gustavo, Carol, Camila. Todos vocês tornaram as minhas 40h semanais em horário comercial menos extenuantes. Vocês dividiram o dia-a-dia comigo e me ajudaram a navegar essas águas tempestuosas.

Ao fim desses agradecimentos, sinto que não foi o suficiente; que lhes ofereço pouco. Mas saibam que o que derramo sobre estas páginas é um infinitésimo diante de tudo que vocês me deram, na forma de carinho, paciência, compreensão e incentivo. Me apoio nos ombros de todos vocês com o objetivo de enxergar mais longe. Obrigado.

"I will take the Ring, though I do not know the way."
J.R.R. Tolkien, The Fellowship of the Ring

Resumo

Hamiltonianos efetivos são um pilar central da óptica quântica e da eletrodinâmica quântica em cavidades. Ao identificar e descartar termos que não contribuem de forma significativa para a dinâmica global em um determinado regime de parâmetros, eles fornecem modelos tratáveis que ainda reproduzem a dinâmica experimentalmente relevante. Em sistemas de interação luz-matéria, um exemplo canônico é a redução da hamiltoniana de Rabi (HR) à hamiltoniana de Jaynes-Cummings (HJC) sob a aproximação de onda girante (AOG), na qual termos que oscilam rapidamente são descartados da hamiltoniana do sistema devido à sua contribuição desprezível para a dinâmica. Embora a seleção de modelos efetivos seja frequentemente obtida de forma analítica, como no caso da AOG, essa abordagem exige um profundo entendimento do sistema físico em questão e, em geral, envolve um elevado grau de intervenção manual. Assim, o problema central abordado nesta dissertação é investigar a possibilidade de inferir uma hamiltoniana efetiva diretamente a partir de dados dinâmicos limitados e de conhecimento físico *a priori* compartilhado por muitos sistemas físicos comuns. Com esse objetivo, empregamos um arcabouço de redes neurais informadas pela física para reproduzir a passagem da HR para a HJC. Nesse contexto, o vetor de estado de valores complexos do sistema composto átomo-cavidade é representado por duas redes neurais de valores reais (correspondentes às partes real e imaginária), treinadas por meio de uma função de perda que: (i) impõe a dinâmica de Schrödinger por meio de um resíduo de equação diferencial ordinária (EDO) avaliado em instantes de tempo fixos; (ii) ajusta observações de dados provenientes de valores esperados experimentalmente acessíveis (número de fótons e população do estado excitado); e (iii) impõe restrições físicas por meio de penalidades associadas às condições iniciais e à normalização do estado. Os parâmetros desconhecidos de acoplamento átomo-cavidade presentes no Hamiltoniano são tratados como parâmetros treináveis da rede neural e aprendidos conjuntamente com a trajetória do estado. Após validar a abordagem em um cenário de modelo compatível (dados simulados com a HJC e uso do resíduo da EDO de Jaynes-Cummings) para diferentes estados iniciais da cavidade, abordamos a identificação de modelos efetivos treinando a rede com uma interação de Rabi decomposta, contendo acoplamentos distintos para os termos girantes e contra-girantes. Os resultados mostram que, em um regime de parâmetros no qual a AOG é válida, o treinamento recupera consistentemente o acoplamento de Jaynes-Cummings enquanto suprime o acoplamento contra-girante, tanto quando os dados são gerados pela dinâmica completa de Rabi quanto quando dados de Jaynes-Cummings são ajustados utilizando a HR mais expressiva. Mostramos ainda que a parametrização escolhida para impor a não negatividade do parâmetro de acoplamento aprendido (função valor absoluto versus função *softplus*) influencia a rapidez com que termos fracamente identificáveis colapsam para zero. Esses resultados sustentam o uso de PINNs inversas como uma abordagem prática para testar se termos candidatos em hamiltonianas completas são de fato suportados pelos dados e para extrair hamiltonianas efetivas a partir de medições parciais de valores esperados.

Palavras-chave: redes neurais informadas pela física; problemas inversos; aprendizado de hamiltonianas; dinâmica quântica; modelo de Rabi; modelo de Jaynes-Cummings; hamiltonaianas efetivas.

Abstract

Effective Hamiltonians are a cornerstone of quantum optics and cavity quantum electro-dynamics (QED). By identifying and discarding terms that do not contribute significantly to the overall dynamics in a given parameter regime, they provide tractable models that still reproduce the experimentally relevant dynamics. In light-matter systems, a canonical example is the reduction of the quantum Rabi Hamiltonian (RH) to the effective Jaynes-Cummings Hamiltonian (JCH) under the rotating-wave approximation (RWA), where fast-oscillating terms are discarded from the system's Hamiltonian due to little contribution to dynamics. While effective-model selection is often justified analytically, as in the RWA, this approach requires a deep understanding of the physical system at hand, and is often heavily handmade. Hence, the central problem addressed in this dissertation is whether one can infer an effective Hamiltonian directly from limited dynamical data and physics a priori knowledge shared by many common physical systems. To this end, we leverage an inverse physics-informed neural network (PINN) framework to reproduce the passage of the RH to the JCH. In this setting, the complex-valued vector state of the composed atom-cavity system is represented by two real-valued neural networks (for the real and imaginary components) trained using a loss function with terms that (i) enforces Schrödinger dynamics through an ordinary differential equation (ODE) residual evaluated at fixed time instants, (ii) matches data observation from experimentally accessible expectation values (photon number and excited-state population), and (iii) imposes physical constraints through initial-condition and normalization penalties. Unknown atom-cavity coupling parameters present in the Hamiltonian are treated as trainable neural network parameters and learned jointly with the state trajectory. After validating the approach in a matched-model baseline (simulating train data with JCH and using the Jaynes-Cummings ODE residual) across distinct cavity initial states, we tackle effective-model identification by training with a split Rabi interaction containing separate rotating and counter-rotating couplings. Results show that in a parameter regime where RWA is feasible, training consistently recovers the Jaynes-Cummings coupling while suppressing the counter-rotating coupling toward zero, both when data are generated by full Rabi dynamics and when Jaynes-Cummings data are fit using the more expressive RH. We further show that the chosen non-negativity parameterization of the learned coupling parameter (absolute value versus softplus functions) influences how readily weakly identifiable terms collapse toward zero. These results support inverse PINNs as a practical workflow for testing whether candidate terms in full Hamiltonians are warranted by data and for extracting effective Hamiltonians from partial expectation-value measurements.

Keywords: physics-informed neural networks; inverse problems; Hamiltonian learning; quantum dynamics; Rabi model; Jaynes-Cummings model; effective Hamiltonians.

List of Figures

List of Tables

List of Acronyms

Adam	Adaptive Moment Estimation
ANN	Artificial Neural Network
JCH	Jaynes-Cummings Hamiltonian
MAE	Mean Absolute Error
MSE	Mean Squared Error
ODE	Ordinary Differential Equation
PDE	Partial Differential Equation
PINN	Physics-Informed Neural Network
QED	Quantum Electrodynamics
QuTiP	Quantum Toolbox in Python
ReLU	Rectified Linear Unit
RH	Rabi Hamiltonian
RWA	Rotating Wave Approximation

Contents

Chapter 1

Introduction

Quantum optics is the area of physics that studies the quantum properties of light and its interaction with matter. By enabling the generation and precise control of nonclassical light (single photons, entanglement, squeezing), quantum optics has been a major engine of modern photonic and quantum technologies, such as quantum key distribution for secure links, the invention of the laser, and squeezed-light injection to boost interferometric sensitivity in gravitational-wave detectors [1, 2, 3]. In this field, the electromagnetic field is modeled as a quantized entity, as opposed to classical descriptions in which light is treated as a continuous wave, and light-matter coupling is described within the formalism of quantum mechanics [4, 5, 6, 7]. Among the models of light-matter interaction is the Rabi model [8], which describes a two-level system coupled to a single quantized field mode. Despite its compact form, obtaining analytic solutions is generally nontrivial, and the use of controlled approximations is often essential [9, 10]. A particularly important approximation is the Rotating Wave Approximation (RWA) [11, 12], in which counter-rotating terms are neglected when their effect on the dynamics is perturbative. Under the RWA, the Rabi model reduces to the Jaynes-Cummings model [11, 13], which is especially relevant in near-resonant and atom-field weak-coupling regimes (i.e. when the atom-field coupling is much slower than the natural atom and field frequencies) widely encountered in cavity QED and related platforms [14, 15]. The Jaynes-Cummings Hamiltonian is therefore naturally interpreted as an effective Hamiltonian capturing the dominant dynamics of the full light-matter interaction within an appropriate subspace and parameter regime [16, 17, 18]. More broadly, effective Hamiltonians are simplified representations of a system's full Hamiltonian designed to preserve the relevant physics (typically low-energy or coarse-grained) while enabling tractable analysis and simulation. In quantum optics and quantum information, effective descriptions routinely arise by eliminating fast or off-resonant degrees of freedom, leading to compact generators of the slow dynamics [19, 20].

Machine learning, on the other hand, is a subfield of Artificial Intelligence concerned with algorithms that learn patterns from data to make predictions, decisions, or to discover structure in

complex datasets [21, 22]. In the last decade, the field has advanced rapidly due to the combined effect of (i) the availability of large-scale datasets and (ii) increased computational power, particularly through specialized hardware and highly optimized software ecosystems, which together enable the training of high-capacity models at scale [22, 23, 24]. As a result, machine learning now impacts a broad range of domains: in computer vision, deep convolutional networks achieved a milestone on large-scale image classification benchmarks [25]; in natural language processing, attention-based architectures (Transformers) became a dominant paradigm for sequence modeling and text generation [26]; in decision-making and control, deep reinforcement learning reached superhuman performance in challenging tasks such as Go [27], a strategy game often used to assess problem solving skills and reasoning of trained AIs; and in scientific and medical applications, deep networks have been successfully applied to medical image diagnosis and to molecular biology problems such as protein structure prediction [28, 29]. Among the most influential model classes in this context are Artificial Neural Networks (ANNs), whose origins trace back to early mathematical abstractions of biological neurons [30] and learning models such as the perceptron [31]. Modern multi-layer neural networks are typically trained by minimizing a loss function via gradient-based optimization, with gradients computed efficiently through the backpropagation algorithm, introduced in early form by [32] and popularized in the modern deep learning era by [33].

In physics, machine learning has become an important tool both for accelerating numerical modeling and for extracting physical information from limited or noisy data. Examples include data-driven closures and surrogate models in fluid dynamics [34, 35, 36] and neural-network-based approaches for solving differential equations [37, 38, 39]. A particularly influential framework is that of Physics-Informed Neural Networks (PINNs) [40], in which governing equations, typically expressed as Ordinary Differential Equations (ODEs) or Partial Differential Equations (PDEs), are incorporated into the training objective, acting as a structured inductive bias or regularization that constrains the hypothesis space toward physically consistent solutions [41, 42, 43]. In addition to solving forward problems, PINNs can be applied to *inverse* problems by learning unknown parameters appearing in the governing equations directly from observations [40, 44]. This capability is particularly appealing for quantum dynamics, where experimentally accessible data often consist of a small set of expectation values rather than full state information. In this context, physics-informed learning offers a route to infer Hamiltonian parameters and, more generally, to support data-driven construction of effective Hamiltonians from more complete microscopic descriptions or experimental measurements [45].

The remainder of this dissertation is organized as follows. In Chapter 2, we motivate the problem and state the main objective and specific research goals. Chapters 3 and 4 introduce the theoretical and methodological foundations used throughout the work: we review the Jaynes-Cummings

and Rabi models (including the rotating-wave approximation), discuss the interaction picture and the role of expectation values, and summarize the neural-network and parameter-learning machinery. Chapter 5 presents the loss construction and training procedure adopted in this study. In Chapter 6, we present and analyze the numerical results: first, the matched-model setting in which both data generation and the PINN residual use the Jaynes-Cummings Hamiltonian (with representative initial conditions such as a Fock superposition and a coherent state), and then the effective-Hamiltonian identification experiments, including both Rabi-simulated data and Jaynes-Cummings-simulated data under a more general training model.

Chapter 2

Motivation and Objectives

A central theme in quantum optics and cavity Quantum Electrodynamics (QED) is that the behavior of a system of interest is rarely determined by an isolated Hamiltonian alone. In realistic scenarios, an "atom" (more generally, a few-level quantum system) is embedded in an electromagnetic environment, interacts with a cavity mode, and may also be coupled to additional degrees of freedom such as external spurious modes or measurement backaction [14, 46, 47]. A standard modeling strategy is therefore to partition the total Hamiltonian as $\hat{H} = \hat{H}_S + \hat{H}_E + \hat{H}_{SE}$, where $\hat{H}_S$ describes the system, $\hat{H}_E$ describes the environment, and $\hat{H}_{SE}$ accounts for their interaction [48, 49]. Even when the environment is traced out and one works with an effective dynamical equation for the reduced state, the resulting description still depends on physical parameters such as transition frequencies, detunings, coupling strengths, and dissipation rates. These parameters determine not only qualitative features (e.g. oscillatory versus overdamped dynamics), but also quantitative predictions required for control, metrology, and device characterization.

Within this broader setting, light-matter interaction models provide a particularly fruitful area to study parameter inference. The quantum Rabi model describes a two-level system coupled to a single quantized cavity mode and is widely used as a minimal yet nontrivial description of cavity and circuit QED [8]. In regimes where the rotating-wave approximation is justified, the Rabi Hamiltonian reduces to the Jaynes-Cummings Hamiltonian [12, 13, 50]. This reduction is more than a mathematical convenience: it exemplifies the broader notion of an *effective Hamiltonian*, namely a simplified version of the full generator of dynamics that accurately reproduces the relevant physics within a restricted parameter regime or subspace [51]. Effective Hamiltonians are essential in quantum physics because they enable tractable analysis and simulation while preserving essential dynamical and spectral properties. In this context, one question that naturally arises is: what is the parameter regime that takes a full Hamiltonian and lands it in an effective Hamiltonian? This question is usually answered by analytical methods such as the RWA that is performed in a full Rabi Hamiltonian to land in the effective Jaynes-Cummings Hamiltonian (see Sec. 3.5).

However, numerical methods to determine these parameter regimes are still a vast field of study. Traditional approaches include spectroscopy-based identification, model fitting via numerical integration, and statistical estimation procedures that require repeated forward simulations [12, 48]. In practice, these approaches may become computationally expensive, particularly when the dynamics is stiff, when the parameter space is high-dimensional, or when the available data are sparse and noisy. Moreover, experimental constraints often mean that one does not have direct access to the full state vector, but only to a subset of expectation values, possibly under partial observability and imperfect initialization.

PINNs offer a conceptually appealing alternative to this parameter estimation effort, known in the field as the inverse problem. The PINN methodology augments a data-misfit objective with a *physics constraint* that penalizes violations of the governing differential equations [40]. In the present context, this governing equation is the Schrödinger equation for state evolution. Because PINNs leverage automatic differentiation to evaluate time derivatives of a neural network surrogate, they can be trained to satisfy the dynamical equations across a continuum of time points while simultaneously matching measured data. Crucially, the unknown physical parameters can be treated as trainable variables and learned together with the neural representation of the solution. This "inverse PINN" perspective turns parameter estimation into a single constrained optimization problem, where the physics acts as a strong inductive bias that can reduce the need for dense and noise-free datasets [40].

Motivated by these considerations, this dissertation investigates whether and how physics-informed learning can be used not only to infer parameters of a chosen Hamiltonian model, but also to connect a *full* Hamiltonian description to a *reduced, effective* Hamiltonian that captures the experimentally relevant dynamics. The Rabi and Jaynes-Cummings models provide a natural starting point: they are physically motivated, experimentally relevant, and they explicitly illustrate the passage from a full model to an effective one via well-informed approximations. Establishing a reliable framework in this setting is a stepping stone toward more realistic scenarios, including open-system dynamics and dissipative processes, where effective descriptions are often indispensable.

Main objective

The main objective of this work is to develop and validate a PINN framework for *Hamiltonian parameter learning* and for the inference of *effective Hamiltonians* from data, with a primary focus on light-matter interaction models, using the Rabi and Jaynes-Cummings Hamiltonians as proof of concept.

Specific objectives

This main objective unfolds into the following specific goals:

- Implement a reproducible simulation and training pipeline for quantum dynamics, enabling controlled generation of training datasets and systematic benchmarking against ground-truth parameters. This simulation and model training framework will be available as open source code.
- Design loss functions that enforce the governing equations of motion in Schrödinger picture, incorporate physical constraints such as normalization of state vectors, and handle complex-valued quantities through consistent real-imaginary decompositions.
- Apply inverse PINN strategies to infer key parameters in the Jaynes-Cummings and Rabi models, such as coupling strengths, using only partial observation channels (e.g., expectation values of experimentally accessible observables).
- Study how learned parameters and learned dynamics relate to effective descriptions, with emphasis on the passage from the Rabi model to the Jaynes-Cummings model and on identifying regimes where an effective Hamiltonian provides a faithful description of the data.
- Produce a modular implementation that facilitates future extensions to more complex models (multi-mode cavities, multi-level emitters, driving terms, and additional interactions), with clearly documented experiments and evaluation protocols.

By pursuing these objectives, the dissertation aims to clarify to what extent physics-informed learning can serve as a principled tool for quantum model discovery and parameter identification, and how it can be used to bridge microscopic Hamiltonians and experimentally meaningful effective descriptions in light-matter interaction systems.

Chapter 3

Jaynes-Cummings and Rabi models

In this chapter, we will expose the physical system of interest for this study: the Rabi model, the most fundamental light-matter interaction model. From now on, unless otherwise stated, all quantities will be expressed in units where $\hbar = 1$.

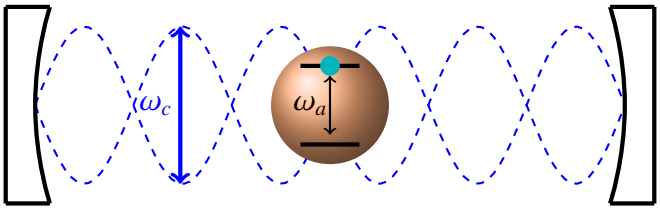


Figure 3.1: Atom-cavity system representation, showing an emitter with energy transition gap ω_a within a cavity with a trapped electromagnetic field of frequency ω_c.

The Jaynes-Cummings and Rabi models are paradigmatic Hamiltonians for light-matter interaction in quantum optics and cavity QED [6, 12, 14, 50]. They describe a two-level emitter (atom, quantum dot, superconducting qubit, ...) coupled to a single quantized mode of the electromagnetic field in a cavity. A schematic representation of this system is shown in Fig. 3.1.

A standard derivation begins by combining three physically motivated reductions [6, 7, 51]: (i) a *two-level truncation* of the matter system; (ii) a *mode expansion* of the quantized electromagnetic field, reduced to a single relevant cavity mode; (iii) the *electric-dipole interaction* between matter and the field. We develop each ingredient carefully, then assemble the Rabi Hamiltonian and show how the Jaynes-Cummings Hamiltonian follows via the RWA.

3.1 Decomposition of the total Hamiltonian

We decompose the Hamiltonian as

$$\hat{H} = \hat{H}_A + \hat{H}_C + \hat{H}_{\text{int}}, \tag{3.1}$$

where $\hat{H}_A$ is the free emitter Hamiltonian, $\hat{H}_C$ is the free cavity-field Hamiltonian, and $\hat{H}_{\text{int}}$ is the interaction term [12, 51]. This separation reflects the physical assumption that, to a good approximation, the cavity and emitter are independently well-defined subsystems, weakly perturbed by their coupling.

3.2 Two-level emitter Hamiltonian

Let $\{|g\rangle, |e\rangle\}$ be two energy eigenstates of the isolated emitter, namely the ground and excited states with eigenenergies E_g and E_e [51, 52], respectively. Restricting the emitter Hilbert space to $\text{span}\{|g\rangle, |e\rangle\}$, the emitter Hamiltonian is

$$\hat{H}_A = E_g \, |g\rangle\langle g| + E_e \, |e\rangle\langle e| \, . \tag{3.2}$$

It is convenient to express this Hamiltonian in terms of Pauli operators. Define the identity in the two-level subspace and $\hat{\sigma}_z$ as

$$\mathbb{I}_a := |g\rangle\langle g| + |e\rangle\langle e| \, , \qquad \hat{\sigma}_z := |e\rangle\langle e| - |g\rangle\langle g| \, . \tag{3.3}$$

Now observe the following algebraic identities:

$$|e\rangle\langle e| = \frac{1}{2} \left(\mathbb{I}_a + \hat{\sigma}_z \right) , \tag{3.4}$$

$$|g\rangle\langle g| = \frac{1}{2} \left(\mathbb{I}_a - \hat{\sigma}_z \right) . \tag{3.5}$$

Substituting these into Eq. (3.2) yield

$$\begin{aligned} \hat{H}_A &= \frac{1}{2} \left(\mathbb{I}_a - \hat{\sigma}_z \right) E_g + \frac{1}{2} \left(\mathbb{I}_a + \hat{\sigma}_z \right) E_e \\ &= \frac{E_e + E_g}{2} \, \mathbb{I}_a + \frac{E_e - E_g}{2} \, \hat{\sigma}_z . \end{aligned} \tag{3.6}$$

The term proportional to $\mathbb{I}_a$ produces only a global phase $e^{-i(E_e+E_g)t/2}$ in Schrödinger evolution and therefore has no effect on observables [51]. Dropping this physically irrelevant energy shift and

defining the transition frequency $\omega_a := E_e - E_g$, we obtain the standard two-level Hamiltonian

$$\hat{H}_A = \frac{\omega_a}{2}\hat{\sigma}_z. \tag{3.7}$$

This will be later used, but it is useful that we also introduce the ladder operators

$$\hat{\sigma}_+ := |e\rangle\langle g|, \qquad \hat{\sigma}_- := |g\rangle\langle e|, \tag{3.8}$$

which satisfy $\hat{\sigma}_+^\dagger = \hat{\sigma}_-$ and $[\hat{\sigma}_+, \hat{\sigma}_-] = \hat{\sigma}_z$ in this two-level subspace.

$$\hat{\sigma}_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \qquad \hat{\sigma}_+ = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad \hat{\sigma}_- = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}. \tag{3.9}$$

3.3 Quantized cavity-field Hamiltonian

Quantization of the electromagnetic field in a cavity proceeds by expanding the field in a complete set of cavity modes (normal modes). Each mode behaves as a quantum harmonic oscillator with annihilation/creation operators $\hat{a}_{\mathbf{k}}/\hat{a}_{\mathbf{k}}^\dagger$ satisfying $[\hat{a}_{\mathbf{k}}, \hat{a}_{\mathbf{k}'}^\dagger] = \delta_{\mathbf{k},\mathbf{k}'}$ [6, 51, 53]. The free-field Hamiltonian reads

$$\hat{H}_{\text{field}} = \sum_{\mathbf{k}} \omega_{\mathbf{k}} \left(\hat{a}_{\mathbf{k}}^\dagger \hat{a}_{\mathbf{k}} + \frac{1}{2} \right), \tag{3.10}$$

where $\omega_{\mathbf{k}}$ is the frequency of the mode $\mathbf{k}$. In cavity QED, one often makes a *single-mode approximation*: if the emitter is near resonance with one cavity mode and far from others, then only that mode contributes appreciably to the dynamics [14, 50]. Denoting this selected mode by $\hat{a}, \hat{a}^\dagger$ with frequency ω_c, we write

$$\hat{H}_C = \omega_c \hat{a}^\dagger \hat{a}. \tag{3.11}$$

As in the emitter case, the constant $\omega_c/2$ is a uniform energy shift and can be dropped without affecting dynamics or expectation values [51]. Since the cavity-mode Hilbert space is spanned by an infinite family of Fock states $\{|n\rangle\}_{n=0}^{\infty}$, numerical implementations require truncating it to a finite subspace $\{|0\rangle, |1\rangle, \ldots, |N\rangle\}$. In this $(N+1)$-dimensional basis, the bosonic operators are represented by the matrices

$$\hat{a} = \begin{pmatrix} 0 & \sqrt{1} & 0 & \cdots & 0 \\ 0 & 0 & \sqrt{2} & \cdots & 0 \\ 0 & 0 & 0 & \ddots & \vdots \\ \vdots & & & \ddots & \sqrt{N} \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix}, \qquad \hat{a}^\dagger = (\hat{a})^T, \tag{3.12}$$

where A^T is the transpose of matrix A and the photon-number operator $\hat{a}^\dagger\hat{a} = \mathrm{diag}(0, 1, \ldots, N)$ is diagonal.

3.4 Dipole coupling and the Rabi Hamiltonian

In the long-wavelength (electric-dipole) approximation, appropriate when the spatial extent of the emitter is much smaller than the wavelength of the relevant field mode, the leading light matter coupling is the dipole interaction [6, 12, 51]

$$\hat{H}_{\mathrm{int}} = -\hat{\mathbf{d}} \cdot \hat{\mathbf{E}}(\mathbf{r}), \tag{3.13}$$

where $\hat{\mathbf{d}}$ is the emitter dipole operator and $\hat{\mathbf{E}}(\mathbf{r})$ is the electric-field operator evaluated at the emitter position $\mathbf{r}$.

To express the dipole operator under our two-level truncation for the emitter, we begin by inserting the two-level system identity $\mathbb{I}_a = |g\rangle\langle g| + |e\rangle\langle e|$ on both sides of $\hat{\mathbf{d}}$:

$$\hat{\mathbf{d}} = \mathbb{I}_a\hat{\mathbf{d}}\mathbb{I}_a = \sum_{i,j\in\{g,e\}} |i\rangle\langle i|\ \hat{\mathbf{d}}\ |j\rangle\langle j| = \mathbf{d}_{gg}\,|g\rangle\langle g| + \mathbf{d}_{ee}\,|e\rangle\langle e| + \mathbf{d}_{ge}\,|g\rangle\langle e| + \mathbf{d}_{eg}\,|e\rangle\langle g| \tag{3.14}$$

with matrix elements $\mathbf{d}_{ij} := \langle i|\,\hat{\mathbf{d}}\,|j\rangle$. For many dipole-allowed transitions, the diagonal elements vanish, $\mathbf{d}_{gg} = \mathbf{d}_{ee} = \mathbf{0}$, while the transition dipole elements $\mathbf{d}_{eg}$ and $\mathbf{d}_{ge} = \mathbf{d}_{eg}^*$ dominate [7, 51]. In this common situation,

$$\hat{\mathbf{d}} = \mathbf{d}_{eg}\,|e\rangle\langle g| + \mathbf{d}_{ge}\,|g\rangle\langle e| = \mathbf{d}_{eg}\hat{\sigma}_+ + \mathbf{d}_{ge}\hat{\sigma}_-. \tag{3.15}$$

Because $|e\rangle$ and $|g\rangle$ are defined only up to arbitrary phase factors, one may choose phases so that the relevant component of $\mathbf{d}_{eg}$ along the cavity polarization is real and positive, simplifying later expressions [51].

Now we turn to the electric field component of Eq. (3.13). For a single cavity mode with unit polarization vector $\boldsymbol{\epsilon}$ and real mode function $u(\mathbf{r})$, the electric field operator takes the form [6, 51, 53]

$$\hat{\mathbf{E}}(\mathbf{r}) = \epsilon\,\mathcal{E}_0\,u(\mathbf{r})\left(\hat{a} + \hat{a}^\dagger\right), \qquad \mathcal{E}_0 := \sqrt{\frac{\omega_c}{2\epsilon_0 V}}, \tag{3.16}$$

where ϵ_0 is the vaccum permittivity constant and V is the effective mode volume (set by cavity geometry and normalization conventions). Placing the emitter at a field antinode is modeled by taking $u(\mathbf{r}) \approx 1$; more generally, $u(\mathbf{r})$ can be absorbed into the definition of the coupling constant.

Substituting the two-level dipole expression and the single-mode field into Eq. (3.13) yields

$$\begin{aligned}\hat{H}_{\text{int}} &= -\left(\mathbf{d}_{eg}\hat{\sigma}_{+} + \mathbf{d}_{ge}\hat{\sigma}_{-}\right)\cdot\left(\boldsymbol{\epsilon}\,\mathcal{E}_0 u(\mathbf{r})(\hat{a}+\hat{a}^{\dagger})\right)\\ &= -\mathcal{E}_0 u(\mathbf{r})\left[(\mathbf{d}_{eg}\cdot\boldsymbol{\epsilon})\hat{\sigma}_{+} + (\mathbf{d}_{ge}\cdot\boldsymbol{\epsilon})\hat{\sigma}_{-}\right](\hat{a}+\hat{a}^{\dagger}).\end{aligned} \tag{3.17}$$

Using $\mathbf{d}_{ge}\cdot\boldsymbol{\epsilon} = (\mathbf{d}_{eg}\cdot\boldsymbol{\epsilon})^{*}$, and choosing phases so this quantity is real, we can write

$$\hat{H}_{\text{int}} = -\mathcal{E}_0 u(\mathbf{r})(\mathbf{d}_{eg}\cdot\boldsymbol{\epsilon})(\hat{\sigma}_{+}+\hat{\sigma}_{-})(\hat{a}+\hat{a}^{\dagger}). \tag{3.18}$$

If we define the (real) coupling strength as

$$g := -(\mathbf{d}_{eg}\cdot\boldsymbol{\epsilon})\,\mathcal{E}_0\,u(\mathbf{r}), \tag{3.19}$$

the interaction then becomes

$$\hat{H}_{\text{int}} = g\,(\hat{a}+\hat{a}^{\dagger})(\hat{\sigma}_{+}+\hat{\sigma}_{-}). \tag{3.20}$$

The overall sign of g is convention-dependent and does not affect measurable physics in this closed-system setting.

Collecting Eqs. (3.7), (3.11), and (3.20) and plugging them in Eq. (3.1) yields the Rabi Hamiltonian [8, 18, 50]

$$\hat{H}_{\text{Rabi}} = \omega_c\,\hat{a}^{\dagger}\hat{a} + \frac{\omega_a}{2}\,\hat{\sigma}_z + g\,(\hat{a}+\hat{a}^{\dagger})(\hat{\sigma}_{+}+\hat{\sigma}_{-}). \tag{3.21}$$

This Hamiltonian contains both energy-conserving exchange processes and so-called *counter-rotating* processes, as made explicit below. Ordering the composite basis as $\{|e,n\rangle\}\cup\{|g,n\rangle\}$ (all excited-atom states followed by all ground-atom states), the Rabi Hamiltonian takes the block form

$$\hat{H}_{\text{Rabi}} = \begin{pmatrix} \omega_c\hat{a}^{\dagger}\hat{a} + \frac{\omega_a}{2}\,\mathbb{I} & g(\hat{a}+\hat{a}^{\dagger}) \\ g(\hat{a}+\hat{a}^{\dagger}) & \omega_c\hat{a}^{\dagger}\hat{a} - \frac{\omega_a}{2}\,\mathbb{I} \end{pmatrix}, \tag{3.22}$$

where $\mathbb{I}$ is the $(N+1)\times(N+1)$ identity. The symmetric off-diagonal blocks, equal to $g(\hat{a}+\hat{a}^{\dagger})$, encode all four coupling terms in Eq. (3.24). For the minimal truncation $N=1$, with basis $\{|e,0\rangle, |e,1\rangle, |g,0\rangle, |g,1\rangle\}$, this yields the explicit 4×4 matrix

$$\hat{H}^{(N=1)}_{\text{Rabi}} = \begin{pmatrix} \frac{\omega_a}{2} & 0 & 0 & g \\ 0 & \omega_c + \frac{\omega_a}{2} & g & 0 \\ 0 & g & -\frac{\omega_a}{2} & 0 \\ g & 0 & 0 & \omega_c - \frac{\omega_a}{2} \end{pmatrix}, \tag{3.23}$$

where quanta-conserving couplings connect $|e,0\rangle \leftrightarrow |g,1\rangle$, while counter-rotating couplings con-

nect $|e,1\rangle \leftrightarrow |g,0\rangle$.

3.5 Rotating-wave approximation and the Jaynes-Cummings Hamiltonian

Expanding the product in Eq. (3.21) results in four terms:

$$g(\hat{a}+\hat{a}^{\dagger})(\hat{\sigma}_{+}+\hat{\sigma}_{-}) = g\left(\hat{a}\hat{\sigma}_{+}+\hat{a}\hat{\sigma}_{-}+\hat{a}^{\dagger}\hat{\sigma}_{+}+\hat{a}^{\dagger}\hat{\sigma}_{-}\right). \tag{3.24}$$

Physically, $\hat{a}\hat{\sigma}_{+}$ describes *absorbing* a photon while *exciting* the atom (energy exchange), while $\hat{a}^{\dagger}\hat{\sigma}_{-}$ describes the reverse (emission). The remaining two terms, $\hat{a}\hat{\sigma}_{-}$ and $\hat{a}^{\dagger}\hat{\sigma}_{+}$, simultaneously lower both subsystems or raise both subsystems, and sometimes are wrongly called *non-physical* since they apparently violate conservation laws. However, Werlang et al. in [54] bring a discussion showing that although this term does not conserve total number of quanta in the system, they still conserve total energy. Hence, it is not a mere mathematical artifact, but a term that has physical meaning.

To justify which terms dominate, we move to the interaction picture with respect to the free Hamiltonian

$$\hat{H}_0 = \omega_c \hat{a}^{\dagger}\hat{a} + \frac{\omega_a}{2}\hat{\sigma}_z, \tag{3.25}$$

and define $\hat{U}_0(t) = e^{-i\hat{H}_0 t}$. Given a generic operator $\hat{O}$, the interaction picture of it is given by

$$\hat{O}_I(t) = \hat{U}_0^{\dagger}(t)\,\hat{O}\,\hat{U}_0(t). \tag{3.26}$$

Using the harmonic-oscillator commutators and the Pauli algebra, one finds [12, 51]

$$\hat{a}_I(t) = \hat{a}\,e^{-i\omega_c t}, \qquad \hat{a}_I^{\dagger}(t) = \hat{a}^{\dagger}e^{+i\omega_c t}, \tag{3.27}$$

and similarly

$$\hat{\sigma}_{+\,I}(t) = \hat{\sigma}_{+}\,e^{+i\omega_a t}, \qquad \hat{\sigma}_{-\,I}(t) = \hat{\sigma}_{-}\,e^{-i\omega_a t}. \tag{3.28}$$

Therefore, each product in Eq. (3.24) acquires an explicit oscillating phase in the interaction pic-

ture:

$$(\hat{a}\hat{\sigma}_+)_I(t) = \hat{a}\hat{\sigma}_+ e^{+i(\omega_a-\omega_c)t}, \tag{3.28a}$$

$$(\hat{a}^\dagger\hat{\sigma}_-)_I(t) = \hat{a}^\dagger\hat{\sigma}_- e^{-i(\omega_a-\omega_c)t}, \tag{3.28b}$$

$$(\hat{a}\hat{\sigma}_-)_I(t) = \hat{a}\hat{\sigma}_- e^{-i(\omega_a+\omega_c)t}, \tag{3.28c}$$

$$(\hat{a}^\dagger\hat{\sigma}_+)_I(t) = \hat{a}^\dagger\hat{\sigma}_+ e^{+i(\omega_a+\omega_c)t}, \tag{3.28d}$$

and the Rabi Hamiltonian in the interaction picture looks like

$$\hat{H}_I^{\text{Rabi}}(t) = g_1\left(\hat{a}^\dagger\hat{\sigma}_- e^{+i(\omega_c-\omega_a)t} + \hat{a}\hat{\sigma}_+ e^{-i(\omega_c-\omega_a)t}\right) + g_2\left(\hat{a}\hat{\sigma}_- e^{-i(\omega_c+\omega_a)t} + \hat{a}^\dagger\hat{\sigma}_+ e^{+i(\omega_c+\omega_a)t}\right). \tag{3.29}$$

The first two terms oscillate at the *detuned* frequency $\omega_a - \omega_c$, while the latter two oscillate at the much larger frequency $\omega_a + \omega_c$. In the weak-coupling regime $g_{1,2} \ll \omega_{a,c}$ and near resonance $|\omega_a - \omega_c| \approx g_{1,2}$, the rapidly oscillating factors $e^{\pm i(\omega_a+\omega_c)t}$ average to approximately zero on the timescale set by the interaction strength, $t \sim 1/g$ [11, 12, 51]. This motivates neglecting the counter-rotating terms $\hat{a}\hat{\sigma}_-$ and $\hat{a}^\dagger\hat{\sigma}_+$, yielding the RWA interaction Hamiltonian

$$\hat{H}^{(\text{RWA})} = g\left(\hat{a}\hat{\sigma}_+ + \hat{a}^\dagger\hat{\sigma}_-\right). \tag{3.30}$$

Substituting Eq. (3.30) into the total Hamiltonian gives the Jaynes-Cummings Hamiltonian

$$\hat{H}_{JC} = \omega_c\,\hat{a}^\dagger\hat{a} + \frac{\omega_a}{2}\,\hat{\sigma}_z + g_1\left(\hat{a}\hat{\sigma}_+ + \hat{a}^\dagger\hat{\sigma}_-\right), \tag{3.31}$$

where in this case $g_1 = g$, with the index used to maintain notation consistency. Eq. (3.31) accurately describes many cavity-QED experiments in the near-resonant, weak-coupling regime [13, 14, 50]. In contrast, when g_1 becomes a non-negligible fraction of ω_a and ω_c (ultra-strong or deep-strong coupling), the counter-rotating terms cannot be neglected and the full Rabi model from Eq. (3.21) is required [18]. In the same block representation as Eq. (3.22), the Jaynes-Cummings Hamiltonian reads

$$\hat{H}_{JC} = \begin{pmatrix} \omega_c\hat{a}^\dagger\hat{a} + \frac{\omega_a}{2}\,\mathbb{I} & g_1\,\hat{a} \\ g_1\,\hat{a}^\dagger & \omega_c\hat{a}^\dagger\hat{a} - \frac{\omega_a}{2}\,\mathbb{I} \end{pmatrix}. \tag{3.32}$$

Comparing with Eq. (3.22), the RWA replaces the symmetric coupling $\hat{a} + \hat{a}^\dagger$ with the asymmetric pair $(\hat{a},\ \hat{a}^\dagger)$. For the same $N = 1$ truncation and basis $\{|e,0\rangle,\ |e,1\rangle,\ |g,0\rangle,\ |g,1\rangle\}$, the JC Hamilto-

nian becomes

$$\hat{H}_{JC}^{(N=1)} = \begin{pmatrix} \frac{\omega_a}{2} & 0 & 0 & g_1 \\ 0 & \omega_c + \frac{\omega_a}{2} & 0 & 0 \\ 0 & 0 & -\frac{\omega_a}{2} & 0 \\ g_1 & 0 & 0 & \omega_c - \frac{\omega_a}{2} \end{pmatrix}. \tag{3.33}$$

Comparing Eq. (3.33) with Eq. (3.23), the effect of the RWA is clear: the counter-rotating entries g at positions $(2,3)$ and $(3,2)$ vanish, so $|e,1\rangle$ and $|g,0\rangle$ become decoupled from each other, leaving only the pair $|e,0\rangle \leftrightarrow |g,1\rangle$ coupled.

Next section addresses a concept that was fundamental when training the neural network to learn the vector state evolution, since it prevents a stiff and undesired behavior.

3.6 Interaction picture and fast phases

In the Schrödinger picture, the state vector $|\psi(t)\rangle$ carries the time dependence while operators representing observables are typically time-independent. The postulate of unitary time evolution is encoded in Schrödinger's equation [51, 55]

$$i\,\frac{\partial}{\partial t}\,|\psi(t)\rangle = \hat{H}(t)\,|\psi(t)\rangle\,, \tag{3.34}$$

where $\hat{H}(t)$ is the possibly time-dependent Hamiltonian. If $\hat{H}$ is time-independent, the solution can be written in terms of the time-evolution operator

$$|\psi(t)\rangle = \hat{U}(t,t_0)\,|\psi(t_0)\rangle\,, \qquad \hat{U}(t,t_0) = e^{-i\hat{H}(t-t_0)}. \tag{3.35}$$

For explicitly time-dependent Hamiltonians, $\hat{U}(t,t_0)$ is still defined as the unitary operator that propagates the state, but its closed-form expression involves time ordering [51]. In all cases, unitarity implies norm conservation,

$$\langle\psi(t)|\psi(t)\rangle = \langle\psi(t_0)|\psi(t_0)\rangle\,, \tag{3.36}$$

and a global phase $|\psi(t)\rangle \mapsto e^{i\phi(t)}\,|\psi(t)\rangle$ does not affect measurable predictions, since they are canceled out (see Sec. 3.7).

Here, the state vector obeys Eq. (3.34), where we can split the Hamiltonian into a "free" part and an interaction part,

$$\hat{H} = \hat{H}_0 + \hat{H}_{\text{int}}, \tag{3.37}$$

which in the Rabi Hamiltonian from Eq. (3.21) can be written as

$$\hat{H}_0 = \omega_c\,\hat{a}^\dagger\hat{a} + \omega_a\,\hat{\sigma}_+\hat{\sigma}_-, \qquad \hat{H}_{\text{int}} = g_1\left(\hat{a}^\dagger\hat{\sigma}_- + \hat{a}\hat{\sigma}_+\right) + g_2\left(\hat{a}\hat{\sigma}_- + \hat{a}^\dagger\hat{\sigma}_+\right). \tag{3.38}$$

When $\omega_{a,c} \gg g_{1,2}$, the complex amplitudes of $|\psi(t)\rangle$ typically carry fast oscillatory phases at frequencies of order ω_a and ω_c, even when physically relevant observables (e.g. populations) evolve on the slower scale set by g_1 (and weak corrections due to g_2). This separation of time scales can make learning the full complex state in the lab frame unnecessarily stiff for gradient-based training, since the fast oscillations are hard for the network to learn.

A standard remedy is to move to the interaction picture with respect to $\hat{H}_0$ [6, 12, 51, 55]. Define the unitary operator

$$\hat{U}_0(t) = e^{-i\hat{H}_0 t}, \tag{3.39}$$

and the interaction-picture state

$$|\tilde{\psi}(t)\rangle = \hat{U}_0^\dagger(t)\,|\psi(t)\rangle = e^{+i\hat{H}_0 t}\,|\psi(t)\rangle\,. \tag{3.40}$$

Differentiating Eq. (3.40) and using the Schrödinger equation yields the interaction-picture dynamics

$$i\,\frac{\partial}{\partial t}\,|\tilde{\psi}(t)\rangle = \hat{H}_I(t)\,|\tilde{\psi}(t)\rangle\,, \tag{3.41}$$

with the interaction-picture Hamiltonian

$$\hat{H}_I(t) = \hat{U}_0^\dagger(t)\hat{H}_{\text{int}}\hat{U}_0(t). \tag{3.42}$$

Eqs. (3.40)–(3.42) show that the fast carrier evolution generated by $\hat{H}_0$ is absorbed into the definition of $|\tilde{\psi}(t)\rangle$, leaving only the interaction to be learned through Eq. (3.41). In practice, when generating simulated training targets, we transform the numerically obtained lab-frame state $|\psi(t)\rangle$ to $|\tilde{\psi}(t)\rangle$ by applying $e^{+i\hat{H}_0 t}$, which removes fast, physically uninformative and component-wise phases from the complex amplitudes.

In our physics-informed neural network, enforcing the Schrödinger equation as loss term can be done in whichever picture is most numerically favorable, as long as the state and Hamiltonian are transformed consistently. Concretely, it means that: (i) if the network outputs $|\tilde{\psi}(t)\rangle$, then the ODE residual must use the interaction-picture Hamiltonian $\hat{H}_I^{\text{Rabi}}(t)$ from Eq. (3.29); (ii) if the network outputs $|\psi(t)\rangle$ in the lab frame, then the residual must use the lab-frame Hamiltonian $\hat{H}$. We adopt the interaction-picture formulation to reduce stiffness and to focus learning capacity on the slow dynamical envelope relevant to parameter estimation in the $\omega_{a,c} \gg g_{1,2}$ regime.

The last section in this chapter introduces a key concept to simulating train data to feed our

neural networks: expectation values, which are regarded as the outcomes of measurements in quantum systems.

3.7 Expectation values

Given an observable represented by a Hermitian operator $\hat{O}$, its expected value at time t is

$$\langle \hat{O} \rangle_t = \langle \psi(t) | \hat{O} | \psi(t) \rangle \,. \tag{3.43}$$

A useful identity connecting Schrödinger equation (Eq. (3.34)) with the time dependence of expectation values is obtained by differentiating Eq. (3.43). For time-independent $\hat{O}$, one finds [51, 55]

$$\frac{\mathrm{d}}{\mathrm{d}t} \langle \hat{O} \rangle_t = i \, \langle [\hat{H}(t), \hat{O}] \, \rangle_t \,, \tag{3.44}$$

and if $\hat{O} = \hat{O}(t)$ has explicit time dependence, an additional term $\left\langle \frac{\partial}{\partial t} \hat{O}(t) \right\rangle_t$ appears. Although we primarily enforce dynamics at the level of the state vector through Eq. (3.34), Eqs. (3.43) and (3.44) justify how observable trajectories are computed from $|\tilde{\psi}(t)\rangle$ and motivate alternative formulations based directly on operator equations of motion.

Chapter 4

Neural Networks and Parameter Learning

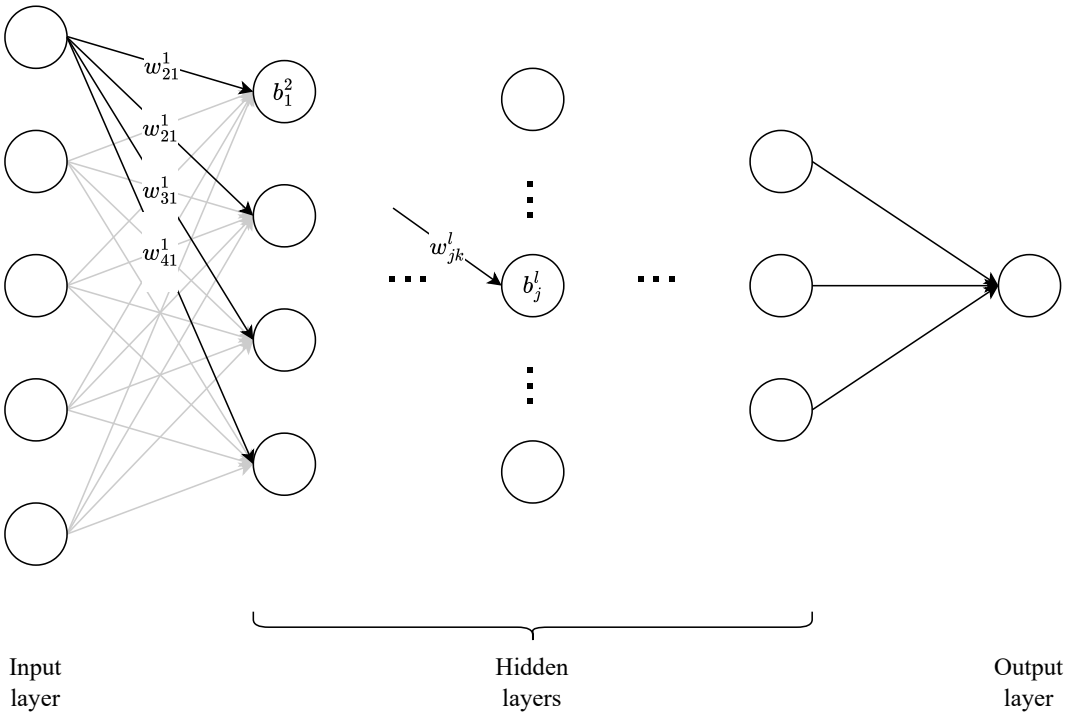


Figure 4.1: Schematic vision of a feed-forward neural network, showing the input layer, as well as hidden layers and an output layer. In this layout, information flows from left to right, passing through the hidden layers to generate an output in the end. w^l_{jk} denotes the weight between neuron k in the layer $l-1$ and neuron j in layer l.

4.1 Artificial Neural Networks

ANNs are parametric function approximators built as compositions of simple nonlinear maps. In the context of this work, we will focus on *feedforward* (fully-connected) networks, where infor-

mation flows from the input to the output through a finite sequence of layers [22, 56]. A schematic view is shown in Fig. 4.1. Throughout, we denote by n_l the number of neurons in layer l, with $l = 0$ the input layer and $l = L$ the output layer.

Let $\mathbf{h}^{l-1} \in \mathbb{R}^{n_{l-1}}$ be the vector of neuron outputs (also called *activations*) from layer $l-1$. A fully-connected layer applies linear combination followed by a nonlinearity:

$$\mathbf{a}^l = \mathbf{W}^l \mathbf{h}^{l-1} + \mathbf{b}^l, \qquad \mathbf{h}^l = \sigma(\mathbf{a}^l), \tag{4.1}$$

where $\mathbf{W}^l \in \mathbb{R}^{n_l \times n_{l-1}}$ is the weight matrix, $\mathbf{b}^l \in \mathbb{R}^{n_l}$ is the bias vector, $\mathbf{a}^l$ is the vector of *pre-activations*, and σ acts component-wise. In index notation, the output of neuron j at layer l is

$$h_j^l = \sigma\left(\sum_{k=1}^{n_{l-1}} w_{jk}^l h_k^{l-1} + b_j^l\right), \tag{4.2}$$

where w_{jk}^l is the weight connecting neuron k in layer $l-1$ to neuron j in layer l [56]. The full network is therefore the composition

$$\tilde{\mathbf{y}}(\mathbf{x}; \mathbf{u}) = f^L \circ f^{L-1} \circ \cdots f^l \circ \cdots \circ f^1(\mathbf{x}), \tag{4.3}$$

where each f^l implements Eq. (4.1), $\mathbf{x} = \mathbf{a}^0$ is the input, $\tilde{\mathbf{y}}$ is the model output, and $\mathbf{u}$ denotes the collection of all trainable parameters $\{\mathbf{W}^l, \mathbf{b}^l\}_{l=1}^L$.

The activation σ is essential: without it, the composition of linear function would collapse into a single linear function, preventing the network from representing nonlinear behavior. Common choices include tanh, sigmoid, ReLU, and sinusoidal activations; the latter often helps when approximating highly oscillatory functions, a behavior frequently encountered in dynamical systems [22, 57] and used throughout this work. More broadly, classical results show that networks with at least one hidden layer and suitable nonlinearities can approximate large classes of continuous functions on compact domains [58], motivating their use as flexible surrogates in physics problems.

4.2 Training as an optimization problem

Given a training set $\mathcal{D} = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^N$, training consists of adjusting the parameters $\mathbf{u}$ from Eq. (4.3) so that $\tilde{\mathbf{y}}(\mathbf{x}_i; \mathbf{u})$ matches $\mathbf{y}_i$ according to a chosen discrepancy measure, usually called loss or cost function and denoted as L [59]. This is formulated as

$$\mathbf{u}^* = \arg\min_{\mathbf{u}} L(\mathbf{u}), \tag{4.4}$$

since $L = L(\mathbf{y}, \tilde{\mathbf{y}}) = L(\mathbf{y}, \tilde{\mathbf{y}}(\mathbf{x}; \mathbf{u}))$ and, given that $\mathcal{D}$ is supposed fixed, then indeed $L = L(\mathbf{u})$. For a regression task, a standard choice is the Mean Squared Error (MSE),

$$L_{\text{data}}(\mathbf{u}) = \frac{1}{N} \sum_{i=1}^{N} \|\tilde{\mathbf{y}}(\mathbf{x}_i; \mathbf{u}) - \mathbf{y}_i\|^2 , \tag{4.5}$$

where $\|\cdot\|$ is the Euclidean norm (for scalar targets this reduces to the usual square). This loss depends on $\mathbf{u}$ only through the network predictions $\tilde{\mathbf{y}}(\cdot; \mathbf{u})$. Another common choice that will be used throughout this project is the Mean Absolute Error (MAE),

$$L_{\text{data}}(\mathbf{u}) = \frac{1}{N} \sum_{i=1}^{N} |\tilde{\mathbf{y}}(\mathbf{x}_i; \mathbf{u}) - \mathbf{y}_i|, \tag{4.6}$$

where the absolute value is computed element-wise.

To minimize $L(\mathbf{u})$, one typically uses gradient-based methods. The simplest update rule is the *gradient descent*,

$$\mathbf{u}_{n+1} = \mathbf{u}_n - \eta \, \boldsymbol{\nabla}_{\mathbf{u}} L(\mathbf{u}_n), \tag{4.7}$$

where $\eta > 0$ is the learning rate. The idea of gradient descent is to update the network's parameters in the direction of most rapidly decrease of the loss function, with step size η. In practice, the gradient $\boldsymbol{\nabla}_{\mathbf{u}} L$ is efficiently computed by the backpropagation algorithm, which applies the chain rule to the layered composition in Eq. (4.3) [22, 56]. Moreover, one usually employs *stochastic* or *mini-batch* variants [22, 60, 61] and adaptive optimizers such as Adam [62, 63]; conceptually, these are refinements of Eq. (4.7) that use noisy estimates of $\boldsymbol{\nabla}_{\mathbf{u}} L$ and/or adapt effective step sizes during training.

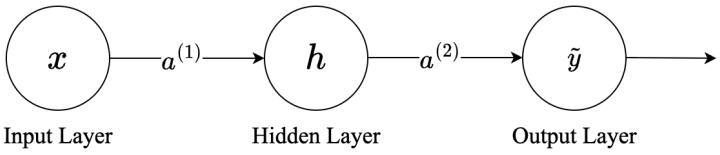


Figure 4.2: A simple neural network setting, with three layers: input, hidden, output, and one neuron per layer. In the picture, x is the input of the neural network; $a^{(1,2)}$ are the linear combinations of the last layer's output with the weights and biases between last neuron and next neuron; h is non-linearity applied to $a^{(1)}$; and $\tilde{y}$ is the non-linearity applied to $a^{(2)}$.

Example 1: Backpropagation algorithm applied to a simple neural network.

Consider the simplest neural network setting, shown in Fig. 4.2, with one input-target pair (x, y), one hidden neuron and one output neuron. Also, consider that the loss function used by this neural network for optimizing its parameters is MSE, and that the activation function is the sigmoid function in all the layers. The intermediary outputs are:

$$a^{(1)} = w_1 x + b_1, \qquad h = \sigma(a^{(1)}),$$
$$a^{(2)} = w_2 h + b_2, \qquad \tilde{y} = \sigma(a^{(2)}),$$

where the sigmoid activation is

$$\sigma(u) = \frac{1}{1+e^{-u}}, \qquad \sigma'(u) = \sigma(u)(1-\sigma(u)), \tag{4.8}$$

and the MSE for a single example is

$$L = \frac{1}{2}(\tilde{y} - y)^2. \tag{4.9}$$

Our primary task is to compute $\partial_{\mathbf{u}} L$, where $\mathbf{u}$ are the weights and biases.
The first chain rule that has to be computed, in order to update the parameters w_2 and b_2, is

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \tilde{y}} \frac{\partial \tilde{y}}{\partial a^{(2)}} \frac{\partial a^{(2)}}{\partial w_2},$$
$$\frac{\partial L}{\partial b_2} = \frac{\partial L}{\partial \tilde{y}} \frac{\partial \tilde{y}}{\partial a^{(2)}} \frac{\partial a^{(2)}}{\partial b_2}.$$

First compute the derivative of the loss with respect to the prediction:

$$\frac{\partial L}{\partial \tilde{y}} = \tilde{y} - y. \tag{4.10}$$

Now, we define the useful quantity δ_2 common to both w_2 and b_2:

$$\delta_2 \equiv \frac{\partial L}{\partial a^{(2)}} = \frac{\partial L}{\partial \tilde{y}} \frac{\partial \tilde{y}}{\partial a^{(2)}} = (\tilde{y} - y)\, \sigma'(a^{(2)}) = (\tilde{y} - y)\, \tilde{y}(1 - \tilde{y}). \tag{4.11}$$

And finally, the derivatives with respect to each parameter,

$$\frac{\partial L}{\partial w_2} = \delta_2 h, \qquad \frac{\partial L}{\partial b_2} = \delta_2. \tag{4.12}$$

Example 1 (continued): Backpropagation algorithm applied to a simple neural network.

Now backpropagate to the hidden neuron to update the pair w_1 and b_1:

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \tilde{y}}\frac{\partial \tilde{y}}{\partial a^{(2)}}\frac{\partial a^{(2)}}{\partial h}\frac{\partial h}{\partial a^{(1)}}\frac{\partial a^{(1)}}{\partial w_1} = \frac{\partial L}{\partial h}\frac{\partial h}{\partial a^{(1)}}\frac{\partial a^{(1)}}{\partial w_1},$$
$$\frac{\partial L}{\partial b_1} = \frac{\partial L}{\partial \tilde{y}}\frac{\partial \tilde{y}}{\partial a^{(2)}}\frac{\partial a^{(2)}}{\partial h}\frac{\partial h}{\partial a^{(1)}}\frac{\partial a^{(1)}}{\partial b_1} = \frac{\partial L}{\partial h}\frac{\partial h}{\partial a^{(1)}}\frac{\partial a^{(1)}}{\partial b_1}.$$

Since $a^{(2)} = w_2 h + b_2$,

$$\frac{\partial L}{\partial h} = \delta_2\, w_2. \tag{4.13}$$

Propagating through the hidden sigmoid $h = \sigma(a^{(1)})$:

$$\delta_1 \equiv \frac{\partial L}{\partial a^{(1)}} = \frac{\partial L}{\partial h}\frac{\partial h}{\partial a^{(1)}} = (\delta_2 w_2)\,\sigma'(a^{(1)}) = (\delta_2 w_2)\,h(1-h). \tag{4.14}$$

Gradients at the hidden layer follow from $a^{(1)} = w_1 x + b_1$:

$$\frac{\partial L}{\partial w_1} = \delta_1\, x, \qquad \frac{\partial L}{\partial b_1} = \delta_1. \tag{4.15}$$

With learning rate $\eta > 0$, one step of gradient descent is:

$$w_2^{(n+1)} = w_2^{(n)} - \eta\,\delta_2^{(n)} h^{(n)}, \qquad b_2^{(n+1)} = b_2^{(n)} - \eta\,\delta_2^{(n)}, \tag{4.16}$$
$$w_1^{(n+1)} = w_1^{(n)} - \eta\,\delta_1^{(n)} x, \qquad b_1^{(n+1)} = b_1^{(n)} - \eta\,\delta_1^{(n)}, \tag{4.17}$$

where n is a generic index representing iteration during training.

While minimizing L_{data} improves agreement with observed data, a purely data-driven approach is agnostic to known physical structure. In many scientific problems, we also know that the target function should satisfy the governing differential equations derived from conservation laws, symmetry principles, or validated phenomenological models. PINNs incorporate this information by adding *equation residual* penalties to the loss, thus biasing training toward physically consistent solutions [40].

4.3 PINNs

In this work we will focus on first-order ordinary differential equations (ODEs) in one independent variable:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{y}(t) = \mathbf{z}(\mathbf{y}(t), t), \tag{4.18}$$

where $\mathbf{y}(t) \in \mathbb{R}^m$ is the state and $\mathbf{z}$ specifies the dynamics. A PINN represents $\mathbf{y}(t)$ by a neural network $\tilde{\mathbf{y}}(t; \mathbf{u})$ that takes t as input and outputs an approximation of the state. A key technical point is that $\frac{\mathrm{d}}{\mathrm{d}t}\tilde{\mathbf{y}}(t; \mathbf{u})$ can be obtained by *automatic differentiation* of the network with respect to its input, which is exact up to floating-point arithmetic and avoids numerical differentiation.

4.3.1 Loss in PINNs

Define the ODE residual

$$\mathbf{r}(t; \mathbf{u}) := \frac{\mathrm{d}}{\mathrm{d}t}\tilde{\mathbf{y}}(t; \mathbf{u}) - \mathbf{z}(\tilde{\mathbf{y}}(t; \mathbf{u}), t). \tag{4.19}$$

The equation-informed loss is constructed by sampling *collocation points* $\{t_k\}_{k=1}^{N_r}$ in the time domain and penalizing $\mathbf{r}$ at these points:

$$L_{\text{eq}} = \frac{1}{N_r} \sum_{k=1}^{N_r} |\mathbf{r}(t = t_k; \mathbf{u})|. \tag{4.20}$$

Importantly, $\{t_k\}$ need not coincide with data measurement times; they are chosen to enforce the physics throughout the domain.

If the initial state $\mathbf{y}(0)$ is known, we impose it via an additional penalty:

$$L_{\text{ic}} = |\tilde{\mathbf{y}}(0; \mathbf{u}) - \mathbf{y}(0)|. \tag{4.21}$$

For more general problems (e.g., PDEs) one would also include boundary conditions, but for the present scope we emphasize initial conditions.

PINN training typically combines (i) a classic data mismatch term L_{data} used in more traditional neural network methods, (ii) the residual loss L_{eq} for the governing equations of the system, (iii) initial condition loss L_{ic}, and possibly (iv) additional regularizers L_{extra} encoding further structure (normalization constraints, symmetries, conserved quantities, etc.). We therefore write

$$L = \alpha_1 L_{\text{data}} + \alpha_2 L_{\text{eq}} + \alpha_3 L_{\text{ic}} + \alpha_4 L_{\text{extra}}, \tag{4.22}$$

where α_ℓ weights the relative importance of each component. In practice, balancing these terms can bring benefits for the training of the neural networks because they can have different numerical scales and converge at different rates. This issue is widely discussed in the PINN literature and motivates strategies such as adaptive reweighting, curriculum schedules, and normalization of residuals [40]. For the scope of this project, we kept $\alpha_\ell = 1, \ \forall \ell$.

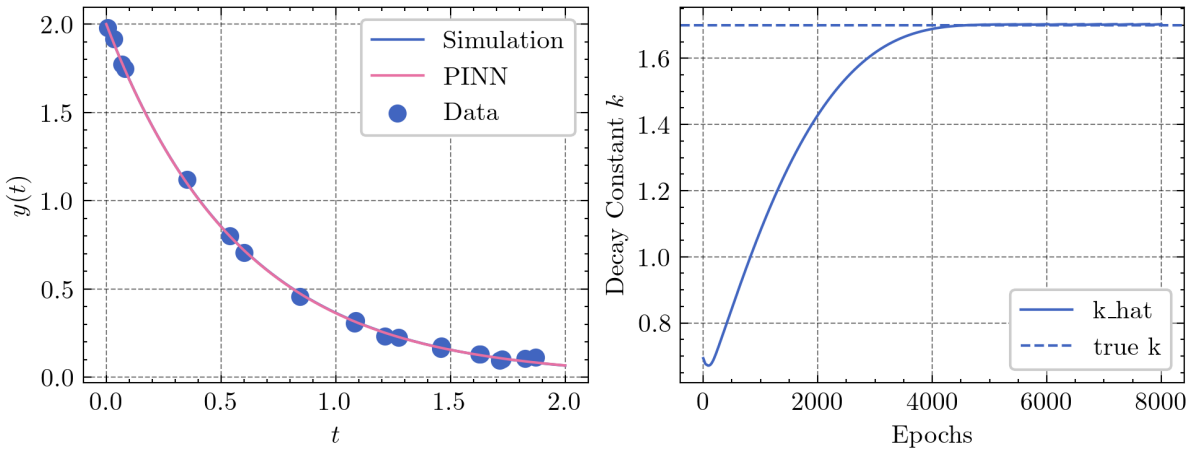


Figure 4.3: (a) The dynamics of the system from Example 2, showing the simulation and the prediction of the neural network superposed, where both are in agreement with the data points. The y-axis shows how the response variable changes with time (for instance it could be the mass of a radioactive sample), while the x-axis is time. (b) Parameter evolution over the training epochs for Example 2. The parameter has a convergence trend towards the real value as training progresses. The y-axis is the parameter value learned by the neural network, and the x-axis shows the epochs of training. The neural network configuration had one neuron on the first layer, 10 neurons on the hidden layer, and 1 neuron on the output layer. Optimizer used was Adam with learning rate of 10^{-3} and the model was trained for 8k epochs.

4.4 Learning unknown parameters (inverse problems)

So far the learning target has been the state function $\mathbf{y}(t)$. A central advantage of PINNs is that they can also infer unknown parameters in the governing equation. Suppose the dynamics depends on a parameter p (or vector $\mathbf{p}$):

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{y}(t) = \mathbf{z}(\mathbf{y}(t), t; p). \tag{4.23}$$

In an *inverse PINN*, we treat p as an additional trainable variable and learn it jointly with $\mathbf{u}$. Concretely, we introduce a learned parameter $\tilde{p}$ and replace Eq. (4.20) by

$$L_{\text{eq}} = \frac{1}{N_r}\sum_{k=1}^{N_r}\left|\frac{\mathrm{d}}{\mathrm{d}t}\tilde{\mathbf{y}}(t = t_k; \mathbf{u}) - \mathbf{z}(\tilde{\mathbf{y}}(t_k; \mathbf{u}), t_k; \tilde{p})\right|, \tag{4.24}$$

so that gradients propagate not only through $\mathbf{u}$ (network weights and biases) but also through $\tilde{p}$. In other words, the optimizer updates $(\mathbf{u}, \tilde{p})$ simultaneously, using the same backpropagation machinery.

Example 2: Exponential decay with parameter identification.

A minimal benchmark to illustrate physics-informed neural networks (PINNs) is the exponential decay model,

$$\frac{\mathrm{d}}{\mathrm{d}t}y(t) + ky(t) = 0, \tag{4.25}$$

where $y(t)$ is the quantity of interest (the mass of some radioactive material sample, for instance) and $k > 0$ is the (radioactive) decay constant. With the initial condition $y(0) = y_0$, the analytical solution is

$$y(t) = y_0 e^{-kt}. \tag{4.26}$$

In this example, $k = 1.7$ and $y_0 = 2.0$.

Let $\tilde{y}(t; \mathbf{u})$ be a neural network approximation of $y(t)$, with parameters $\boldsymbol{u}$. In addition, we introduce $\tilde{k}$ as a *trainable scalar* (to be learned together with $\boldsymbol{u}$). Defining the ODE residual,

$$r(t; \mathbf{u}, \tilde{k}) := \frac{\mathrm{d}}{\mathrm{d}t}\tilde{y}(t; \mathbf{u}) + \tilde{k}\tilde{y}(t; \mathbf{u}), \tag{4.27}$$

we can derive the three terms discussed previously:

1. *EDO loss*: Given N_f collocation points $\{t_j^{(f)}\}_{j=1}^{N_f}$, we impose the differential equation by minimizing

$$L_{\mathrm{eq}}(\mathbf{u}, \tilde{k}) = \frac{1}{N_f}\sum_{j=1}^{N_f}\left[r(t = t_j; \mathbf{u}, \tilde{k})\right]^2. \tag{4.28}$$

2. *Initial condition loss*: We choose the initial condition $y(0) = y_0$, enforced by

$$L_{\mathrm{ic}}(\mathbf{u}) = \left[\tilde{y}(0; \mathbf{u}) - y_0\right]^2. \tag{4.29}$$

3. *Data loss*: Assume we have N_d measurements $\{(t_i^{(d)}, y_i)\}_{i=1}^{N_d}$. The supervised component is

$$L_{\mathrm{data}}(\mathbf{u}) = \frac{1}{N_d}\sum_{i=1}^{N_d}\left[\tilde{y}_i(t_i; \mathbf{u}) - y_i\right]^2. \tag{4.30}$$

A typical weighted combination is

$$L(\theta, \hat{k}) = \alpha L_{\mathrm{data}} + \beta L_{\mathrm{eq}} + \gamma L_{\mathrm{ic}}. \tag{4.31}$$

In this example, $\alpha = \beta = \gamma = 1$. The key point is that $\tilde{k}$ appears explicitly inside the residual $r(t; \mathbf{u}, \tilde{k})$ in Eq. (4.27), and therefore inside $L_{\mathrm{eq}}(\mathbf{u}, \tilde{k})$. During training, backpropagation com-

Example 2 (continued): Exponential decay with parameter identification.

putes gradients $\partial L / \partial \tilde{k}$ together with $\partial L / \partial \mathbf{u}$, so an optimizer (e.g., Adam) updates $\tilde{k}$ in the same loop as the network weights. Intuitively: if $\tilde{k}$ is too large or too small, the residual cannot be made small across collocation points without distorting $\tilde{y}(t)$, so optimization pushes $\tilde{k}$ toward the value that best reconciles the data, the ODE, and the initial condition.

Fig. **??** shows how the model learned the dynamics of the system, and Fig. **??** shows how the model learned the decaying parameter. The neural network used had a simple configuration, with one hidden layer containing 10 neurons and input/output layers both have only one neuron. Optimizer used was Adam with learning rate of 10^{-3} and the model was trained for 8k epochs.

4.4.1 Inductive bias for the parameter learned

It is worth noting that depending on the prior knowledge about the parameter to be learned, an extra inductive bias can be used on how we parameterize the parameter to be learned. For instance, at the system investigated in Example 2, we know before hand that, since it is an exponential decay equation, the parameter k must be strictly positive. In practice, enforcing this constraint could be implemented by adding a learnable parameter θ in the neural network and passing it through an inductive bias function ξ, so that $k = \xi(\theta)$, where ξ could be, in the example discussed above, the absolute value function, or other functions such as the softplus function, a smooth and continuously differentiable alternative to the ReLU [64].

Chapter 5

Methods in the present study

In this work, the dynamics of the atom-cavity system are represented by the state vector $|\psi(t)\rangle \in \mathbb{C}^a$, where a denotes the dimension of the truncated Hilbert space. For a two-level emitter and a cavity Fock-space truncation at mode $n_{\max}$, one has $a = 2(n_{\max} + 1)$. Our neural network is trained to approximate the map

$$t \longmapsto |\tilde{\psi}(t; \mathbf{u})\rangle \in \mathbb{C}^a,$$

where $\mathbf{u}$ collects the trainable parameters of the network (and, in the inverse setting, also the trainable Hamiltonian parameters). In practice, we evaluate the network at a set of time instants $\{t_k\}_{k=1}^N$, yielding predictions that can be arranged as an $a \times N$ array.

5.1 Loss data

Following the PINN framework [40], we identify the terms in the total loss from Eq. (4.22). The data term L_{data} is constructed from simulated expectation values at time instants $\{t_k\}$. In the present study we generate these targets using QuTiP [65], and focus on two experimentally meaningful observables: the cavity photon number $\hat{a}^\dagger \hat{a}$ and the excited-state population projector $\hat{\sigma}_+ \hat{\sigma}_-$, for a L_{data} that reads

$$L_{\text{data}} = \frac{1}{2} \sum_{\hat{O} \in \{\hat{a}^\dagger \hat{a}, \hat{\sigma}_+ \hat{\sigma}_-\}} \frac{1}{N} \sum_{k=1}^{N} \left| \langle \tilde{\psi}(t_k) | \hat{O} | \tilde{\psi}(t_k) \rangle - \langle \psi(t_k) | \hat{O} | \psi(t_k) \rangle \right|, \tag{5.1}$$

where the sum is computed for all the observables being considered. One thing to notice is that the second term in the sum, $\langle \psi(t_k) | \hat{O} | \psi(t_k) \rangle$, although suggesting a calculated term, is actually an experiment measurement.

5.2 Loss initial conditions

The initial-condition loss L_{ic} encodes prior knowledge about the system preparation at $t = 0$. In the closed-system simulations considered here, we specify an initial pure state $|\psi(0)\rangle$ and propagate it under the chosen Hamiltonian. In general, for an atom-cavity system we take a product preparation

$$|\psi(0)\rangle = |\phi_a(0)\rangle \otimes |\phi_c(0)\rangle\,, \tag{5.2}$$

where $|\phi_a(0)\rangle \in \mathbb{C}^2$ is the atomic initial state and $|\phi_c(0)\rangle$ is the cavity-field initial state in the truncated Fock space. While $|\phi_a(0)\rangle$ may be fixed (e.g. $|g\rangle$ or $|e\rangle$) depending on the experiment under study, we varied $|\phi_c(0)\rangle$ across qualitatively different families of photonic states to assess the robustness of the network and of the inverse learning procedure.

Fock (number) state. The simplest cavity preparation is a number state $|n\rangle$, defined as an eigenstate of the number operator $\hat{a}^\dagger\hat{a}$,

$$\hat{a}^\dagger\hat{a}\,|n\rangle = n\,|n\rangle\,, \qquad n \in \mathbb{N}, \tag{5.3}$$

with $\{|n\rangle\}_{n=0}^{\infty}$ forming an orthonormal basis for the single-mode field [53, 12].

Superposition of Fock states. To probe sensitivity to phase and interference effects, we also considered coherent superpositions of number states,

$$|\phi_c(0)\rangle = \sum_{n=0}^{n_{\max}} c_n\,|n\rangle\,, \qquad \sum_{n=0}^{n_{\max}} |c_n|^2 = 1, \tag{5.4}$$

where the complex coefficients c_n set the initial photon-number distribution and relative phases. Such states are particularly useful for testing whether the PINN can faithfully learn dynamics beyond photon-number populations, since phase information can influence observable trajectories through interference [7].

Coherent state. A coherent state $|\alpha\rangle$ is defined as an eigenstate of the annihilation operator,

$$\hat{a}\,|\alpha\rangle = \alpha\,|\alpha\rangle\,, \qquad \alpha \in \mathbb{C}, \tag{5.5}$$

and admits the Fock-basis expansion

$$|\alpha\rangle = e^{-|\alpha|^2/2} \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}}\,|n\rangle\,. \tag{5.6}$$

Coherent states are minimum-uncertainty states whose photon-number statistics are Poissonian with mean $\langle \hat{a}^\dagger \hat{a} \rangle = |\alpha|^2$, and they provide a natural "quasi-classical" benchmark for cavity dynamics [66, 12].

5.3 Loss norm

Since quantum vector states should be normalized at all times [51], that is, given a quantum state vector $|\psi(t)\rangle$, it is always true that $|\langle \psi(t)|\psi(t)\rangle|^2 = 1, \ \forall t$. This fact motivates the normalization loss to have the form

$$L_{\text{norm}} = \frac{1}{N} \sum_{k=1}^{N} \Big| \langle \tilde{\psi}(t_k)|\tilde{\psi}(t_k)\rangle^2 - 1 \Big|. \tag{5.7}$$

5.4 Loss ODE

The differential-equation term L_{eq} is built from the Schrödinger Eq. (3.34). Because both $|\psi(t)\rangle$ and $\hat{H}$ are generally complex, we write $|\tilde{\psi}\rangle = |\tilde{\psi}_R\rangle + i\,|\tilde{\psi}_I\rangle$ and $\hat{H} = \hat{H}_R + i\hat{H}_I$ (this notation is not to be confused with the interaction picture Hamiltonian H_I presented in Sec. 3.6, Eq. (3.42)), with real and imaginary parts taken entrywise. Substituting these decompositions into Schrödinger's equation (Eq. (3.34)) and equating real and imaginary parts yields the coupled real system

$$\begin{aligned} \frac{\partial}{\partial t} |\tilde{\psi}_R(t)\rangle - \hat{H}_R\,|\tilde{\psi}_I(t)\rangle - \hat{H}_I\,|\tilde{\psi}_R(t)\rangle = 0, \\ \frac{\partial}{\partial t} |\tilde{\psi}_I(t)\rangle + \hat{H}_R\,|\tilde{\psi}_R(t)\rangle - \hat{H}_I\,|\tilde{\psi}_I(t)\rangle = 0. \end{aligned}$$

These residual equations are enforced in the loss by evaluating the time-derivatives through automatic differentiation and penalizing the residual norm at the collocation points $\{t_k\}_{k=1}^N$:

$$L_{\text{eq}}^R = \frac{1}{N} \sum_{k=1}^{N} \left| \frac{\mathrm{d}}{\mathrm{d}t} |\tilde{\psi}_R(t=t_k)\rangle - \hat{H}_R\,|\tilde{\psi}_I(t_k)\rangle - \hat{H}_I\,|\tilde{\psi}_R(t_k)\rangle \right|, \tag{5.7a}$$

$$L_{\text{eq}}^I = \frac{1}{N} \sum_{k=1}^{N} \left| \frac{\mathrm{d}}{\mathrm{d}t} |\tilde{\psi}_I(t=t_k)\rangle + \hat{H}_R\,|\tilde{\psi}_R(t_k)\rangle - \hat{H}_I\,|\tilde{\psi}_I(t_k)\rangle \right|. \tag{5.7b}$$

The residual loss L_{eq} depends on the Hamiltonian inserted into Schrödinger's equation. For this study we consider two settings.

(i) *Jaynes-Cummings ODE loss.* In the baseline scenario, the Jaynes-Cummings Hamiltonian is both used to simulate train data for the neural network and for the ODE loss term. The learnable parameter is the Jaynes-Cummings coupling strength, denoted as g_1, which enters

the residual Schrödinger equation.

(ii) *Rabi ODE loss.* In this setting, the ODE loss term is the Rabi Hamiltonian in the interaction picture, as in Eq. (3.29). For this effective Hamiltonian discovery, we chose two parameter settings for simulating the data that will feed the neural network. They differ only on the coupling strengths g_1 and g_2.

In the first setting, we consider $\omega_a = \omega_c = 100$ and $g_1 = g_2 = 1$, corresponding to data simulated under full Rabi dynamics. In this regime, where $\omega_{a,c} \gg g_{1,2}$, we expect to show that the neural network is able to identify that the counter-rotating terms are effectively negligible and converge to the Jaynes-Cummings Hamiltonian.

The second setting uses $\omega_a = \omega_c = 100$, $g_1 = 1$, and $g_2 = 0$, which corresponds to data simulated under the Jaynes-Cummings dynamics. This experiment supports the idea that when the data are simulated using a Jaynes-Cummings dynamics, the inverse-PINN procedure can infer that the counter-rotating component is not supported by the data and correctly learns the Jaynes-Cummings parameters.

In addition, these experiments explore the role of inductive bias in parameter identification by comparing two different positivity-enforcing transformations for $g_{1,2}$: the absolute value and the softplus function. That is, we test the scenarios where $g_{1,2} = |\theta_{1,2}|$ and $g_{1,2} = \text{softplus}(\theta_{1,2})$, according to the discussing at the end of Sec. 4.4.

All the code used to simulate, train the neural networks, and plotting the figures used throughout this study can be found in this repository.

Chapter 6

Results and Analysis

This section presents the main numerical results obtained with the PINN framework developed in this work and discusses them in the context of the specific objectives stated in Chapter 2.

Throughout, we work in the interaction picture and in resonance (as defined in Sec. 3.6), and we use the dimensionless time gt in plots whenever convenient. The PINN training objective is to reconstruct the (complex) state $|\psi(t)\rangle$ via two real-valued networks, $|\tilde{\psi}_R(t)\rangle$ and $|\tilde{\psi}_I(t)\rangle$, and to simultaneously learn the coupling parameters appearing in the residual ODE. Model performance is assessed at three complementary levels: (i) convergence of the composite loss and its components (data, residual ODE, normalization, and initial-condition terms); (ii) agreement between predicted and simulated state amplitudes; and (iii) agreement between predicted and simulated observables (expected values).

In all the experiments shown, unless otherwise stated, the inductive bias used for the coupling parameters was $g = |\theta|$, where θ is the learnable parameter created in the neural network.

6.1 Jaynes-Cummings ODE loss

In this setting, we establish a baseline for the experiments, where the Jaynes-Cummings Hamiltonian is used both for simulating the training data and for the ODE term composing the loss function. These are simpler experiments conceptually, since the neural network only has to learn one parameter; however, since this is a baseline, we tested multiple initial conditions to assess how robust it is to different dynamics and to higher dimensions of the composed system. The parameters used in this set of experiments, including data simulation, neural network settings, and training process, are shown in Table 6.1.

Table 6.1: Parameters used in the simulation, neural network, and training for the experiments simulating data with the Jaynes-Cummings Hamiltonian and using the Jaynes-Cummings Hamiltonian in the L_{eq} term, with different cavity initial conditions. Repeated values across experiments are omitted.

		Fock state	Fock superposition	Coherent
Simulation	Atom initial state	$\|e\rangle \otimes \|0\rangle$	$\|g\rangle \otimes \frac{\|0\rangle+\|1\rangle+\|2\rangle}{\sqrt{3}}$	$\|g\rangle \otimes \|\alpha\rangle$, $\alpha = 1$
	Coupling parameter		$g = 1$	
	Frequency parameters		$\omega_a = \omega_c = 100$	
	Hamiltonian frame		interaction	
	Atom dimension		2	
	Cavity dimension	2	3	10
	Simulation duration		2π	
	Number of data samples	200	200	1000
Network	Neurons in hidden layers		[50, 50, 50]	
	Activation function		sin	
	Dropout probability		0	
	Number of extra parameters		1	
Training	Number of epochs	30k	30k	50k
	Number of points in L_{data}	200	200	500
	Optimizer		Adam	
	Loss		MAE	
	Learning rate		0.001	
	Bias on coupling		absolute value	

6.1.1 Fock initial condition

We proceed to show the results for simulations made using other initial states when the task of the neural network is to learn the coupling parameter g_1 using the Jaynes-Cummings Hamiltonian in the L_{eq} term of the loss function, as in Eq (3.31).

Figure 6.1 presents the evolution of the loss function during the neural network training, showing the individual loss components as well as the total loss L. All loss terms exhibit a pronounced decrease during the initial stages of training, followed by convergence to a stable regime that persists until the end of the optimization. Notably, the L_{eq} term consistently converges to a higher floor compared to the remaining loss components, a behavior that recurs across all the subsequent experiments. This effect can be attributed to multiple factors: first, L_{eq} is evaluated in a multi-dimensional setting, whereas L_{data} is one-dimensional; second, the differential equation residual is

computed over many time instants, effectively increasing the number of contributions in Eqs. (5.8); and finally, L_{eq} explicitly incorporates the atom-cavity coupling parameters $g_{1/2}$, such that the regularization associated with these parameters is accounted for within this loss term.

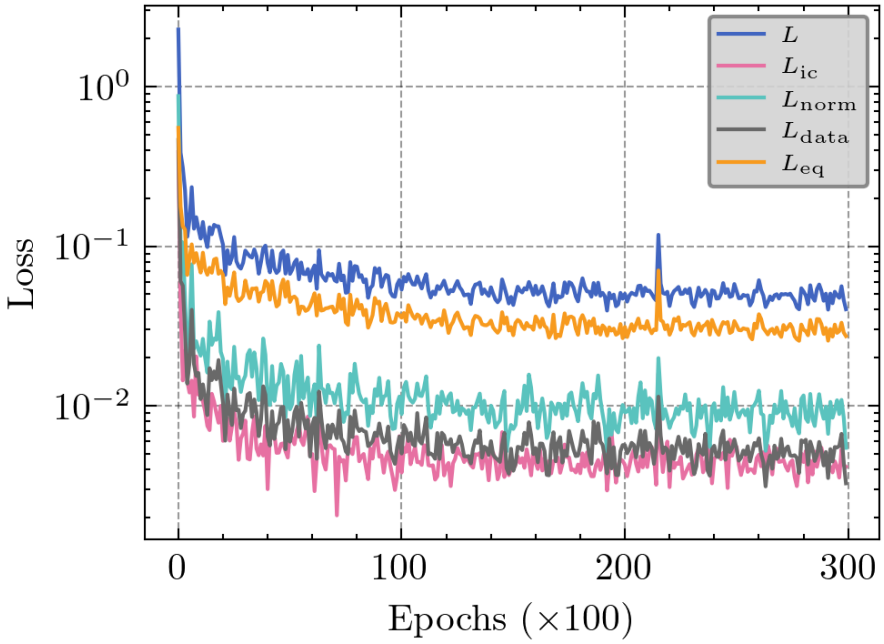


Figure 6.1: Total loss and its components discriminated during the training phase of the Jaynes-Cummings data simulation and the Jaynes-Cummings loss ODE experiment, initial condition is $|e\rangle \otimes |0\rangle$. All the components show an accentuated drop at the beginning of the training process and steadily converge to a lower value. An interesting effect observed here is almost synchronous behavior of all the loss trajectories, showing good balance in the optimization of all components of the total loss function.

Fig. 6.2 shows the reconstruction of the vector state of the system by the neural network. The plots vertically show the basis vectors $|s, n\rangle$ (s for the atom and n for the cavity), and horizontally show scaled time gt. The picture overall shows a good agreement between the predicted states (plots on the top) and simulated states (plots in the middle). The good agreement can be numerically verified by the bottom row of the figure, which shows the absolute error between predictions and simulations, with an error $\sim 10^{-3}$.

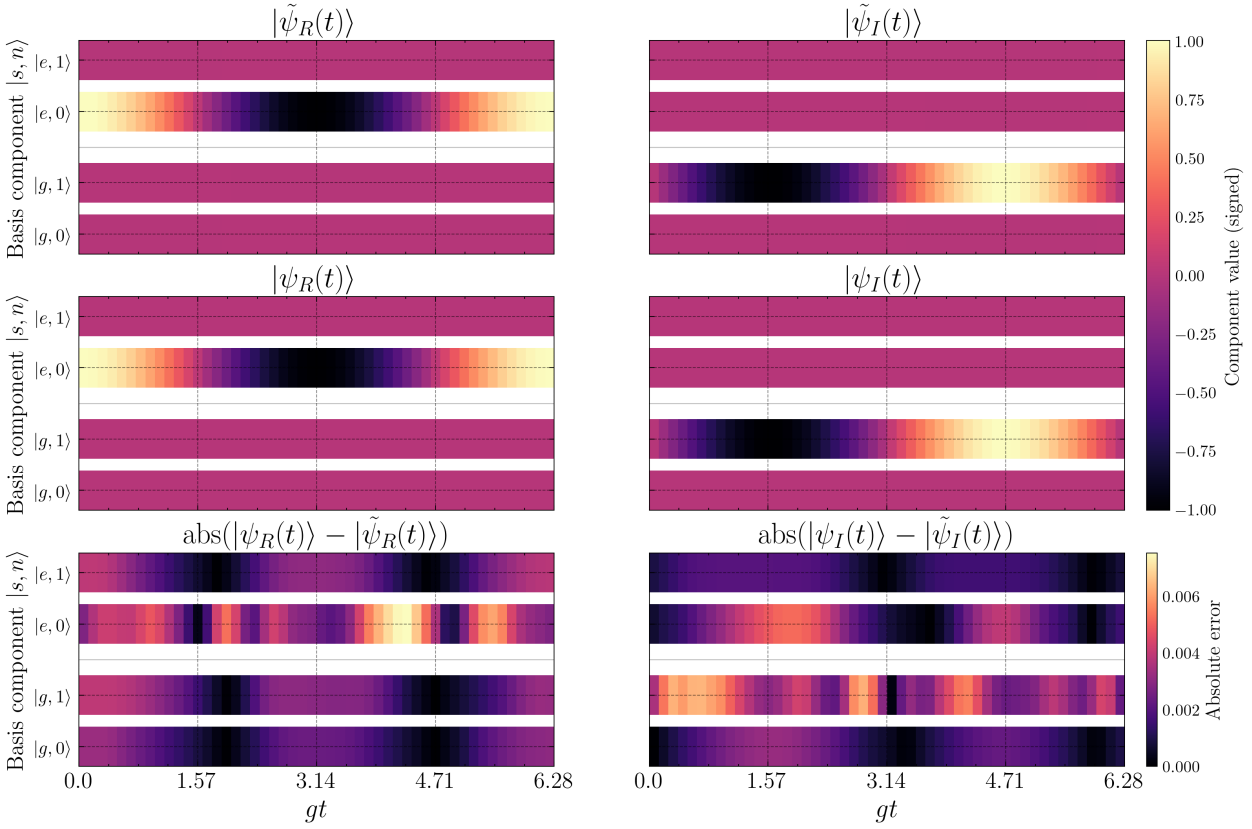


Figure 6.2: State reconstruction in the Jaynes-Cummings data simulation and the Jaynes-Cummings loss ODE experiment, initial condition is $|e\rangle \otimes |0\rangle$. The y-axis shows the basis vectors $|s,n\rangle$ (s is associated with the atom and n with the cavity) of the composed system, while the x-axis shows the time evolution in the characteristic gt time scale. Top rows: learned real/imaginary amplitudes. Middle rows: simulated (target) amplitudes. Bottom rows: absolute difference, element-wise. In this setting, we see a strong agreement of the learned and simulated data, evidenced by the bottom row of the plot, where the absolute error scale is around 10^{-3}.

To summarize state agreement over time with a scalar diagnostic, we compute the fidelity $\mathcal{F}(|\tilde{\psi}(t)\rangle, |\psi(t)\rangle) := \left|\langle\tilde{\psi}(t)|\psi(t)\rangle\right|^2$, which measures how much the simulated and learned states overlap.[1] The fidelity overall stays close to unit, which indicate a good overlap of the predictions of the neural network and simulations.

[1] When both states involved in the fidelity calculation are physical, the result should not be bigger than unity, since states must be normalized at all times. Since this is not a hard constraint in our neural networks, this numerical artifact can be seen in the fidelity plots throughout.

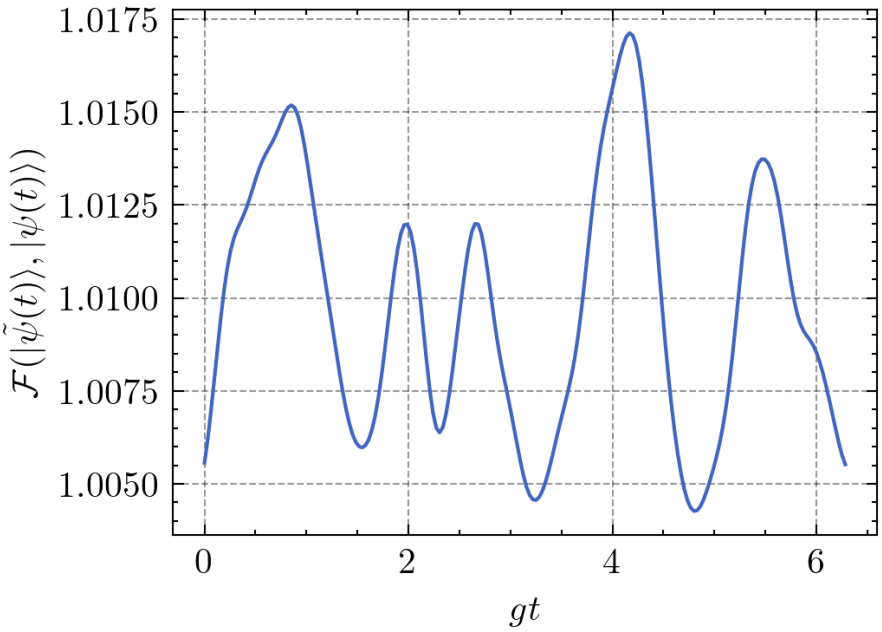


Figure 6.3: Fidelity of the learned state in the Jaynes-Cummings data simulation and the Jaynes-Cummings loss ODE experiment, with initial condition $|e\rangle \otimes |0\rangle$. The values remain close to unity across time, showing that the predicted and simulated states overlap. At some time intervals, we can see that $\mathcal{F} > 1$, which happens because normalization is not a hard constraint in the neural network and small deviations where $\left|\langle\tilde{\psi}(t)|\tilde{\psi}(t)\rangle\right|^2 > 1$ can happen.

Fig. 6.4 compares the learned and simulated expectation values for $\langle\hat{a}^\dagger\hat{a}\rangle$ and $\langle\hat{\sigma}_+\hat{\sigma}_-\rangle$, with the error in the box below the plot defined as

$$\text{error}(\hat{O}) := \left|\langle\tilde{\psi}(t)|\hat{O}|\tilde{\psi}(t)\rangle - \langle\psi(t)|\hat{O}|\psi(t)\rangle\right|.$$

The figure shows a high precision between prediction and simulation, with errors at the 10^{-2} scale throughout the simulation time. The good fit with observables show that the neural network is not only learning a physical behavior of the system enforced by the L_{eq} loss term (and that could be fitted simply with the trivial solution to the Schrödinger equation), but also matches the data.

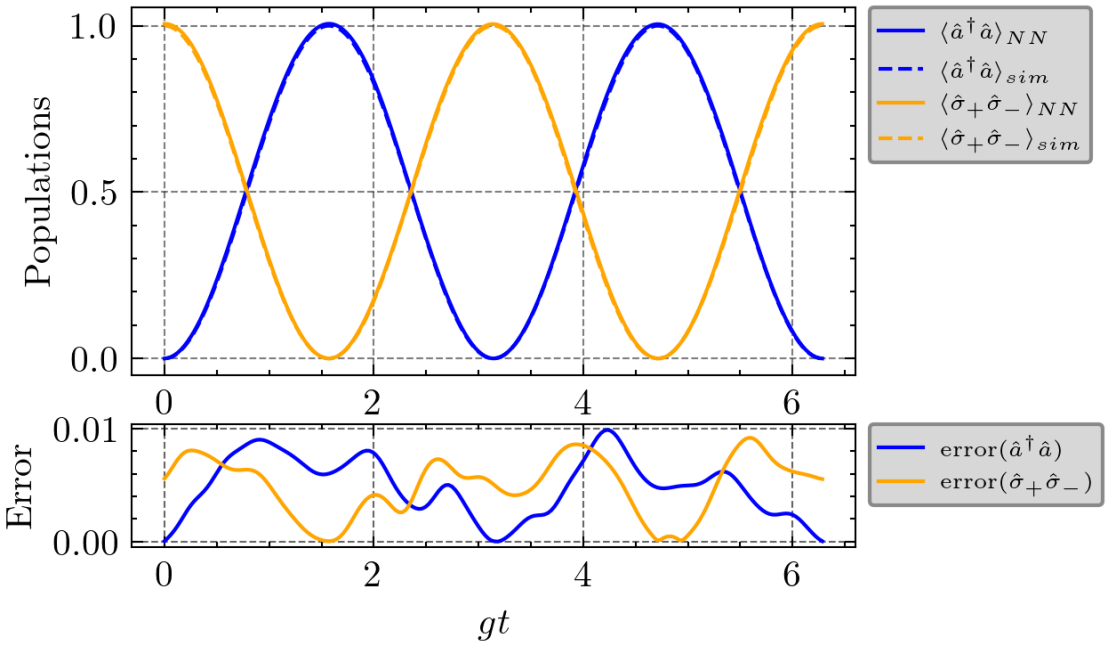


Figure 6.4: Expected values of $\hat{a}^\dagger \hat{a}$ (regarding the cavity) and $\hat{\sigma}_+ \hat{\sigma}_-$ (regarding the atom) operators in the Jaynes-Cummings data simulation and the Jaynes-Cummings loss ODE experiment, with initial condition $|e\rangle \otimes |0\rangle$. Solid lines: PINN predictions computed from $|\tilde{\psi}(t)\rangle$. Dashed lines: simulated targets. The bottom panel reports the signed error $\langle \tilde{\psi}(t)|\hat{O}|\tilde{\psi}(t)\rangle - \langle \psi(t)|\hat{O}|\psi(t)\rangle$ for each observable, remaining small over the full interval.

Fig. 6.5 shows the learning trajectory of the coupling parameter g_1 over the training epochs. We can see a rapid convergence of the parameter in the first few thousand epochs, and then it fluctuates around the ground truth stably. This term is updated through the ODE loss term, so when it converges to the ground truth it also contributes to the overall lowering of L_{eq} term.

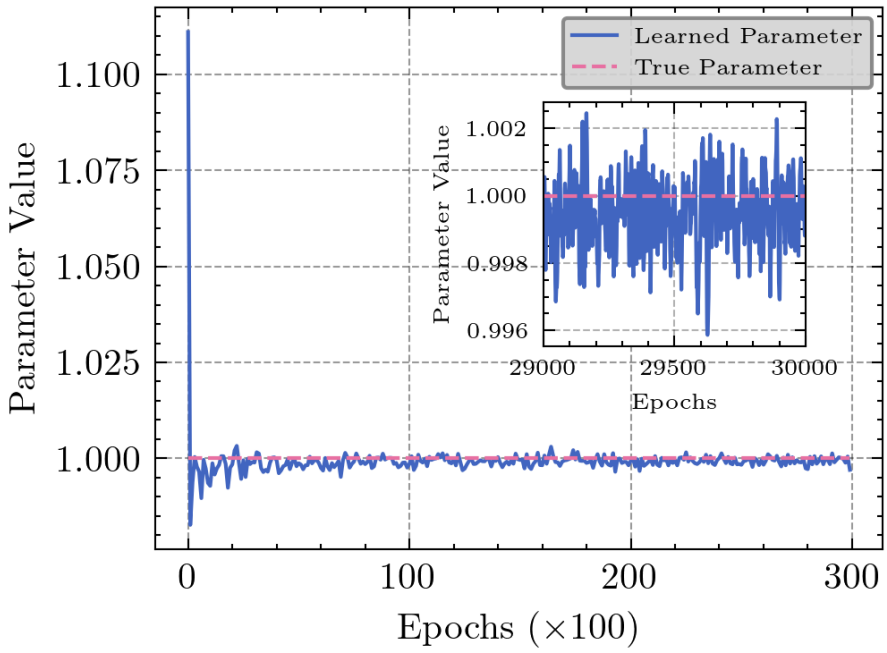


Figure 6.5: Coupling parameter learning, g_1, in the Jaynes-Cummings data simulation and Jaynes-Cummings loss ODE experiment, with initial condition $|e\rangle\otimes|0\rangle$. The learned coupling converges to the ground-truth value and stabilizes, supporting the objective of inferring Hamiltonian parameters from dynamical data. The inset shows the convergence of the parameter in the last 1000 epochs of train.

6.1.2 Fock superposition initial condition

Figs. 6.6–6.10 summarize the training behavior and the quality of the reconstruction for the Fock-superposition initial condition.

Fig. 6.6 shows the evolution of the total loss and its components. After a fast initial transient where all terms terms drop some orders of magnitude, they all remain decreasing more stably around their respective lower floors, indicating that the optimization reaches a regime where the network can simultaneously satisfy the initial condition, keep the state normalized, match the observable data, and reduce the Schrödinger residual. Importantly, the ODE-residual term does not collapse at the expense of the data term (or vice-versa), suggesting that the loss balancing used is sufficient to prevent the training from converging to a purely data-interpolating solution that violates the dynamics, or to a purely dynamics-satisfying solution that ignores the data (that usually converges to the trivial solution).

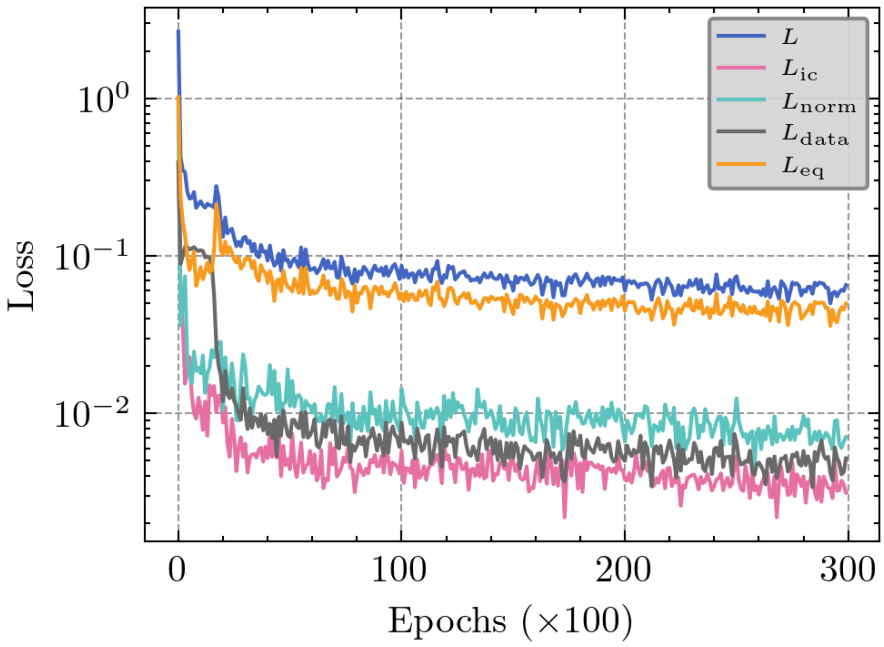


Figure 6.6: Total loss L and its components discriminated during the training phase of the Jaynes-Cummings data simulation and Jaynes-Cummings loss ODE experiment, initial condition is $|g\rangle \otimes \frac{1}{\sqrt{3}} \sum_{k=0}^{2} |k\rangle$. The total loss and its components decrease and stabilize, indicating a consistent compromise between data fitting and physics enforcement. All the terms in the loss function have the same unitary weight.

Fig. 6.7 compares the learned and simulated state amplitudes (real and imaginary parts) as functions of time and basis $|s, n\rangle$ where s concerns the atom and n the cavity, and also reports the absolute error. The learned amplitudes reproduce the qualitative structure of the simulated dynamics across the full time window, while the error map highlights that the remaining discrepancy is localized in time and remains small compared to the typical amplitude scale. This agreement is consistent with the intended role of the ODE residual: it regularizes the learned trajectory so that the network cannot arbitrarily match the observables without reconstructing a physically consistent state.

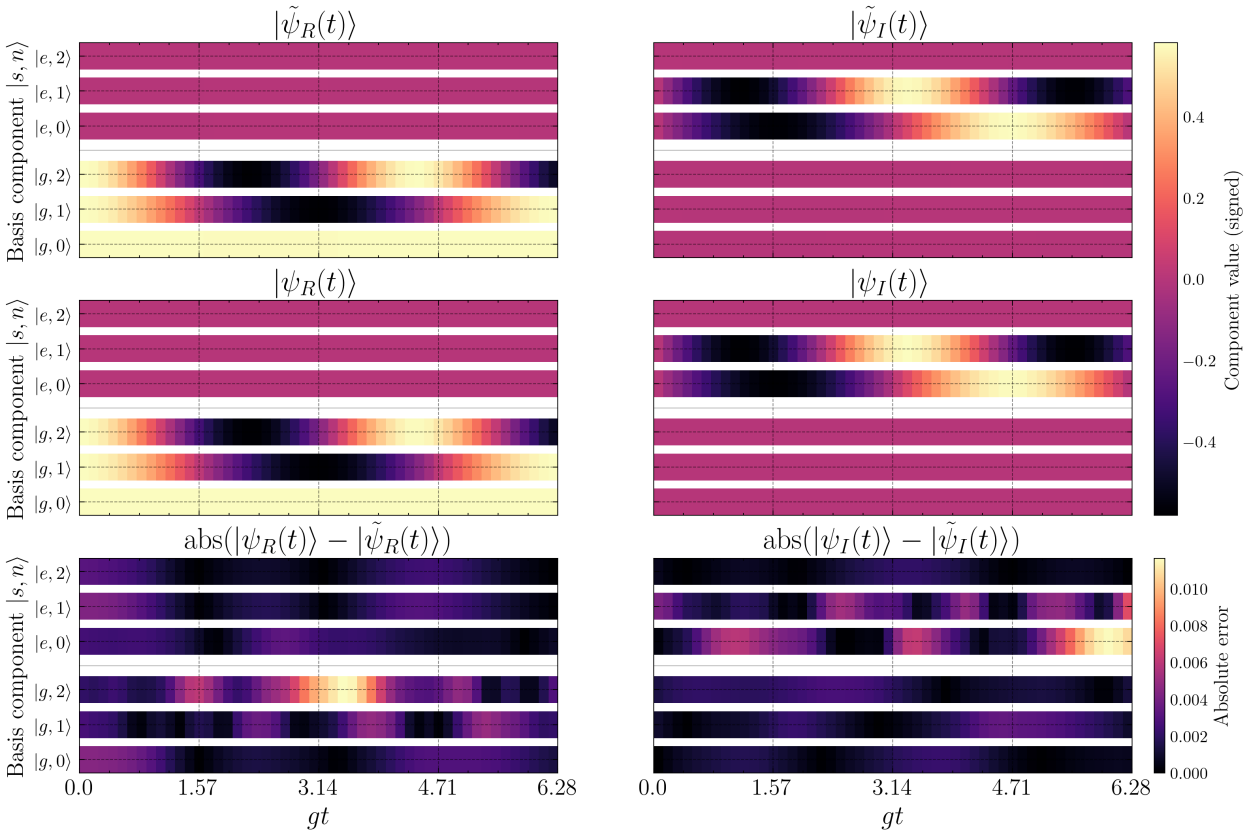


Figure 6.7: State reconstruction in the Jaynes-Cummings data simulation and Jaynes-Cummings loss ODE experiment, initial condition is $|g\rangle \otimes \frac{1}{\sqrt{3}} \sum_{k=0}^{2} |k\rangle$. The y-axis shows the basis vectors $|s,n\rangle$ (s is associated with the atom and n with the cavity) of the composed system, while the x-axis shows the time evolution in the characteristic gt time scale. Top rows: learned real/imaginary amplitudes. Middle rows: simulated (target) amplitudes. Bottom rows: absolute difference, element-wise. The color bar in the top two rows represent the numeric value of the components, and in the bottom row represents the error magnitude. The reconstructed dynamics track the simulated state across the full interval, with residual error remaining small and structured.

Fig. 6.8 displays the fidelity evolution during the time of simulation. Its values remain close to unity over the full time window, indicating high similarity between predicted and simulated states. Small excursions around 1 can arise from finite-precision effects and/or slight deviations from exact normalization when reconstructing the complex state from the two real networks; in practice, the key observation is that the curve stays near 1 with small relative variation, which is consistent with the small error maps in Fig. 6.7.

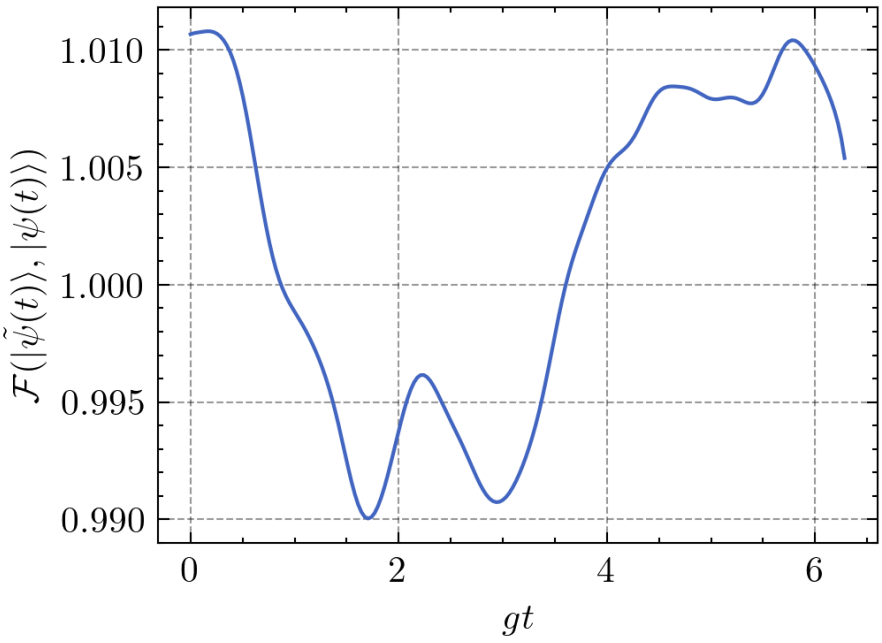


Figure 6.8: Fidelity of the learned state in the Jaynes-Cummings data simulation and Jaynes-Cummings loss ODE experiment, with initial condition $|g\rangle \otimes \frac{1}{\sqrt{3}} \sum_{k=0}^{2} |k\rangle$. The curve remains close to unity throughout the time window, indicating strong agreement between learned and simulated states. At some time intervals, we can see that $\mathcal{F} > 1$, which happens because normalization is not a hard constraint in the neural network and small deviations where $\left| \langle \tilde{\psi}(t) | \tilde{\psi}(t) \rangle \right|^2 > 1$ can happen.

Fig. 6.9 shows the expected values computed using the learned states versus the ones computed using simulated data. Since these observables are the quantities directly constrained by the data loss, their agreement validates that the learned state is not only close in amplitude space but also reproduces physically relevant measurements. The associated error curves remain small compared to the observable scale, which supports the objective of predicting relevant observables over time once the state is reconstructed.

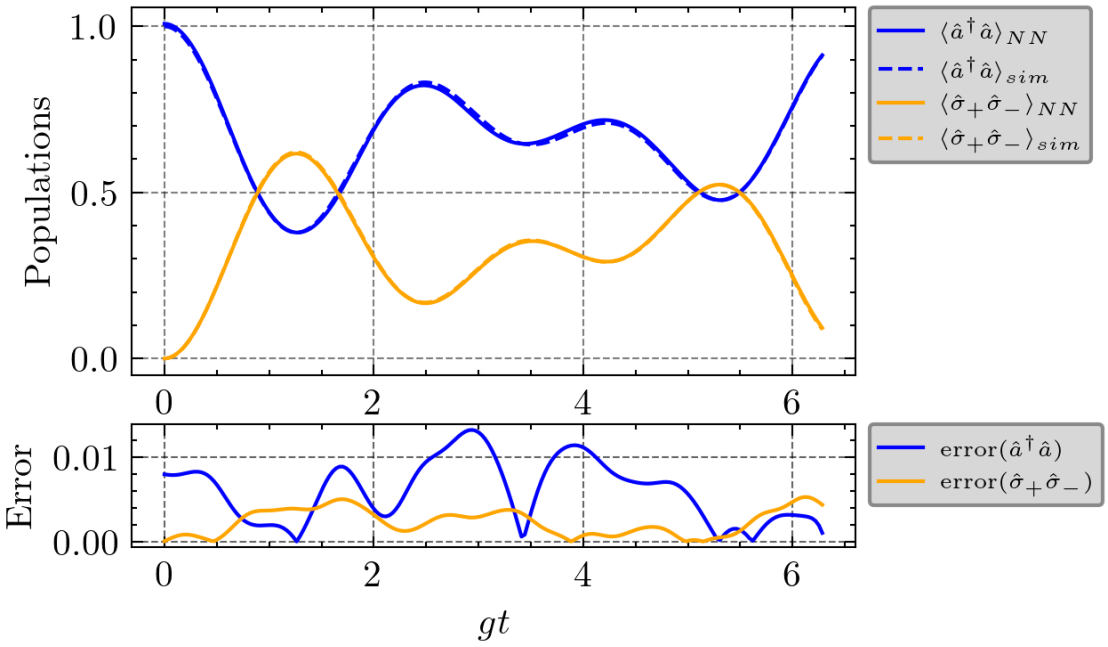


Figure 6.9: Expected values of $\hat{a}^\dagger \hat{a}$ (regarding the cavity) and $\hat{\sigma}_+ \hat{\sigma}_-$ (regarding the atom) operators in the Jaynes-Cummings data simulation and Jaynes-Cummings loss ODE experiment, with initial condition $|g\rangle \otimes \frac{1}{\sqrt{3}} \sum_{k=0}^{2} |k\rangle$. Solid lines: PINN predictions computed from $|\tilde{\psi}(t)\rangle$. Dashed lines: simulated targets. The bottom panel reports the signed error $\left| \langle \tilde{\psi}(t)|\hat{O}|\tilde{\psi}(t)\rangle - \langle \psi(t)|\hat{O}|\psi(t)\rangle \right|$ for each observable, remaining small over the full interval.

Finally, Fig. 6.10 shows the evolution of the learned coupling parameter g_1. After a brief transient, g_1 converges rapidly to the ground-truth value used in simulation and remains stable, indicating that the inverse-PINN formulation is able to identify the correct Hamiltonian parameter in this matched-model scenario, with the Jaynes-Cummings Hamiltonian being used both in data simulation and in the L_{eq} term.

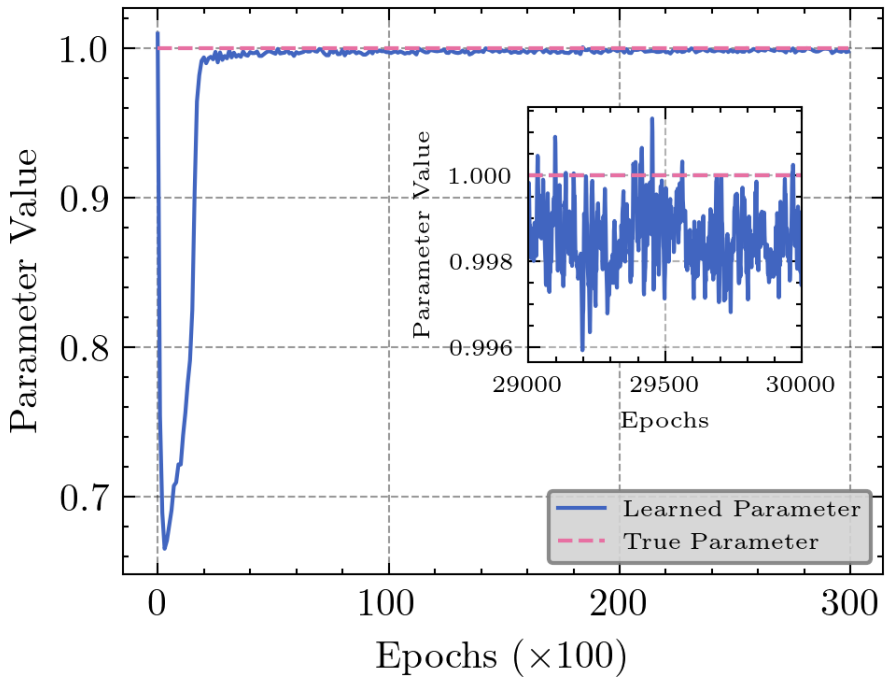


Figure 6.10: Coupling parameter learning, g_1, in the Jaynes-Cummings data simulation and Jaynes-Cummings loss ODE experiment, with initial condition $|g\rangle \otimes \frac{1}{\sqrt{3}} \sum_{k=0}^{2} |k\rangle$. The inset shows the convergence of the parameter in the last 1000 epochs of train. The learned coupling converges to the ground-truth value and stabilizes, supporting the objective of inferring Hamiltonian parameters from dynamical data.

6.1.3 Coherent-state initial condition

We next consider an initial cavity coherent state, which constitutes a qualitatively different regime from the few-state superposition cases. Coherent states populate multiple photon-number levels simultaneously, so accurately reconstructing $|\psi(t)\rangle$ requires the network to capture a broader distribution of amplitudes (including components with comparatively small weight) while still matching the measured observables and respecting the Jaynes-Cummings dynamics through the ODE residual.

Fig. 6.11 shows that the total loss decreases steadily and remains stable throughout training, with sub-losses contributing in a balanced manner. We can also observe some plateaus in the initial condition and the data losses, consistent with the increased difficulty of simultaneously fitting multi-dimensional cavity subsystem and enforcing the data constraint.

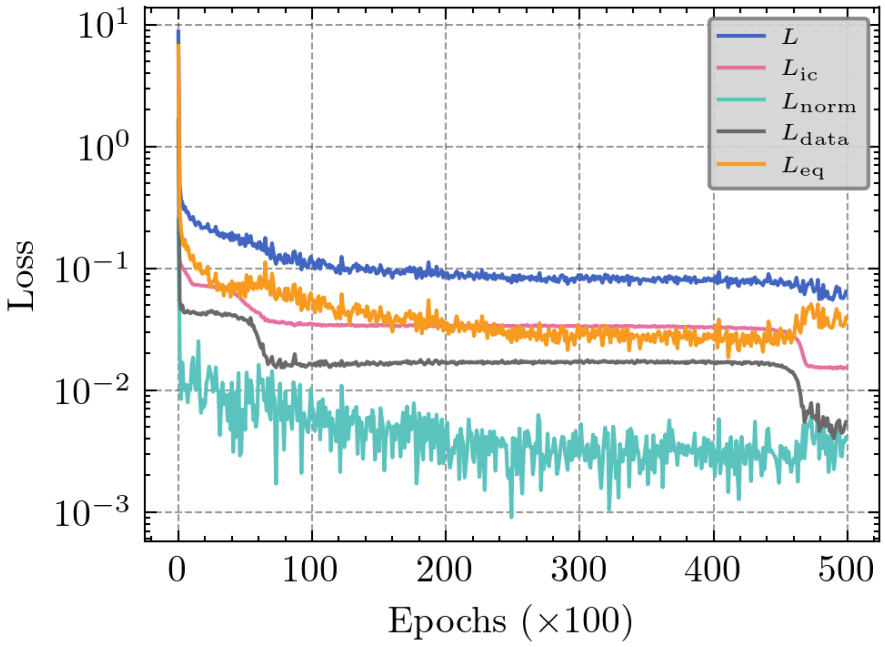


Figure 6.11: Total loss and its components discriminated during the training phase of the Jaynes-Cummings data simulation and Jaynes-Cummings loss ODE experiment, initial condition is $|g\rangle \otimes |\alpha\rangle$, where $|\alpha\rangle = e^{-|\alpha|/2} \sum_{k=0}^{10} \frac{\alpha^k}{\sqrt{n!}} |k\rangle$, $\alpha = 1$. All components decrease and stabilize, where we can see two plateaus on the L_{ic} and L_{data} terms that drop to a lower level at the end of the training loop. The weights in the terms of the total loss function are all equal to unit.

The learned and simulated state amplitudes are compared in Fig. 6.12. The reconstructed real and imaginary parts reproduce the main dynamical features across the full time interval, including the temporal modulation of the dominant basis components. The absolute-error maps show localized regions with larger discrepancies (with peaks at the level of 10^{-2} to 10^{-1} in the color scale), which is expected in coherent-state dynamics: small-amplitude components can be relatively harder to fit while contributing weakly to the primary observables, as can be seen in components $|g, 4\rangle$ and $|g, 5\rangle$ from $|\tilde{\psi}_R(t)\rangle$, as well as in components $|e, 3\rangle$ and $|e, 4\rangle$ from $|\tilde{\psi}_I(t)\rangle$. The error plots indicate that the neural network outputs are close to the trivial solution at these components, since they display a structured oscillatory behavior.

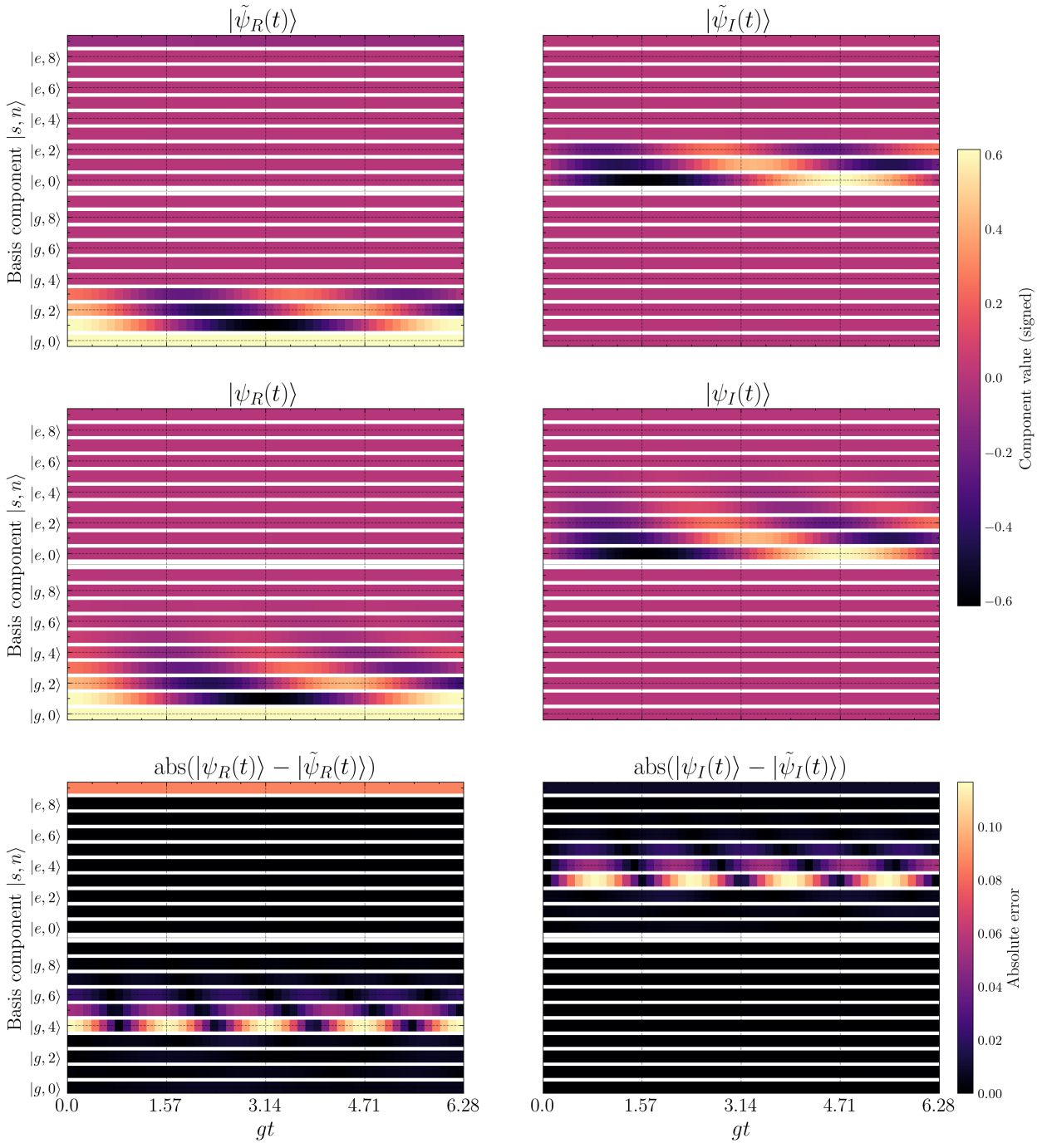


Figure 6.12: State reconstruction in the Jaynes-Cummings data simulation and Jaynes-Cummings loss ODE experiment. Initial condition is $|g\rangle \otimes |\alpha\rangle$, where $|\alpha\rangle = e^{-|\alpha|/2} \sum_{k=0}^{10} \frac{\alpha^k}{\sqrt{n!}} |k\rangle$, $\alpha = 1$. The y-axis shows the basis vectors of the composed system $|s,n\rangle$, s is associated with the atom and n with the cavity, (the basis index only shows even components of the cavity for clarity of the plot), while the x-axis shows the time evolution in the characteristic gt scale. Top rows: learned real/imaginary amplitudes. Middle rows: simulated (target) amplitudes. Bottom rows: absolute difference, element-wise. The color bars represent the number value shown in the main plots. The network captures the global dynamical structure, with remaining discrepancies localized at specific components of the composed state vector.

The fidelity $\mathcal{F}(|\tilde{\psi}(t)\rangle, |\psi(t)\rangle)$ is reported in Fig. 6.13. The curve stays close to unity over the entire window, supporting that the learned trajectory remains highly consistent with the simulated state despite the broader support induced by the coherent initial condition.

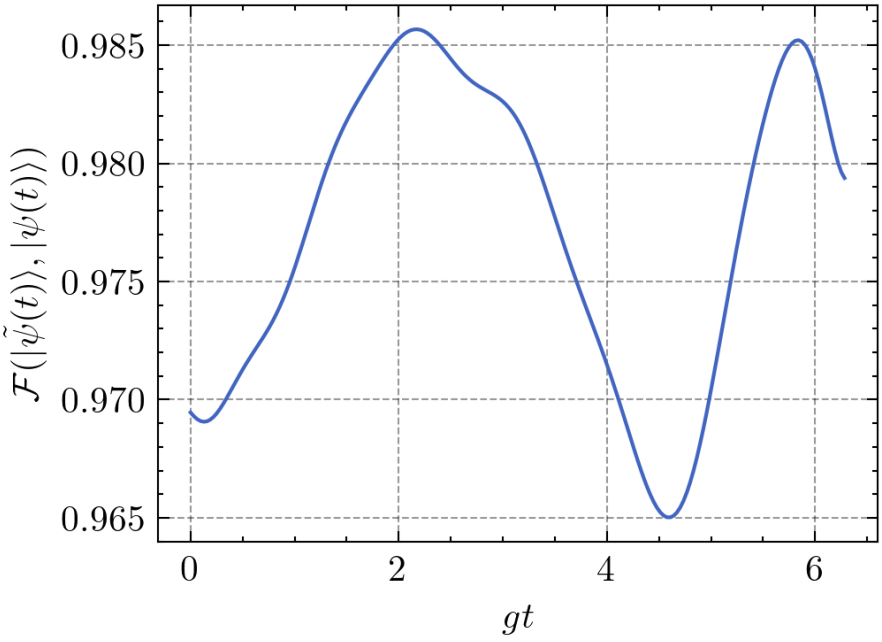


Figure 6.13: Fidelity of the learned state in the Jaynes-Cummings data simulation and Jaynes-Cummings loss ODE experiment. Initial condition $|g\rangle \otimes |\alpha\rangle$, $\alpha = 1$. The values remain close to unity across time, indicating strong state-level agreement.

Fig. 6.14 shows the predicted and simulated expectation values $\langle \hat{a}^\dagger \hat{a} \rangle$ and $\langle \hat{\sigma}_+ \hat{\sigma}_- \rangle$. The PINN closely tracks the target curves, and the reported errors remain small compared to the observable scale (on the order of 10^{-3} to 10^{-2}), confirming that the learned state reproduces the measurement data used in training.

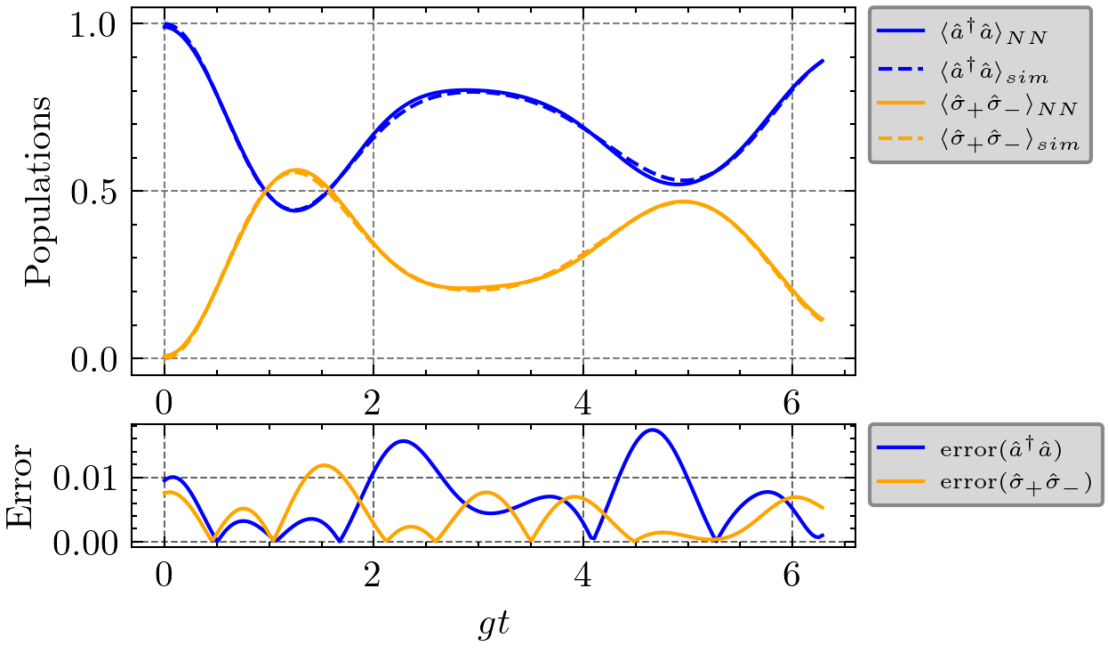


Figure 6.14: Expected values of $\hat{a}^\dagger \hat{a}$ (regarding the cavity) and $\hat{\sigma}_+ \hat{\sigma}_-$ (regarding the atom) operators in the Jaynes-Cummings data simulation and Jaynes-Cummings loss ODE experiment, with initial condition $|g\rangle \otimes |\alpha\rangle$, where $|\alpha\rangle = e^{-|\alpha|/2} \sum_{k=0}^{10} \frac{\alpha^k}{\sqrt{n!}} |k\rangle$, $\alpha = 1$. Solid lines: PINN predictions computed from $|\tilde{\psi}(t)\rangle$. Dashed lines: simulated targets. The bottom panel reports the signed error $\left| \langle \tilde{\psi}(t)|\hat{O}|\tilde{\psi}(t)\rangle - \langle \psi(t)|\hat{O}|\psi(t)\rangle \right|$ for each observable, remaining small over the full interval.

Fig. 6.15 reports the learned coupling g_1. After the characteristic initial transient, g_1 converges toward the ground truth throughout the majority training epochs, remaining stable until almost the end of training, where it encounters a rapid drop around 46k epochs of training, followed by an amplitude correction of a few orders of magnitude. Analyzing Fig. 6.11, this drop in the learned parameter coincides with the increase of the L_{eq} term, implying a worse governing equation matching. However, the total loss term experiences a slight drop because the terms L_{data} and L_{ic} drop due to better data matching.

This behavior indicates that even when the initial cavity state spreads weight over multiple number states, the inverse-PINN formulation can still identify the Jaynes-Cummings coupling consistently and resolve deviations made throughout the training process to converge to the ground truth.

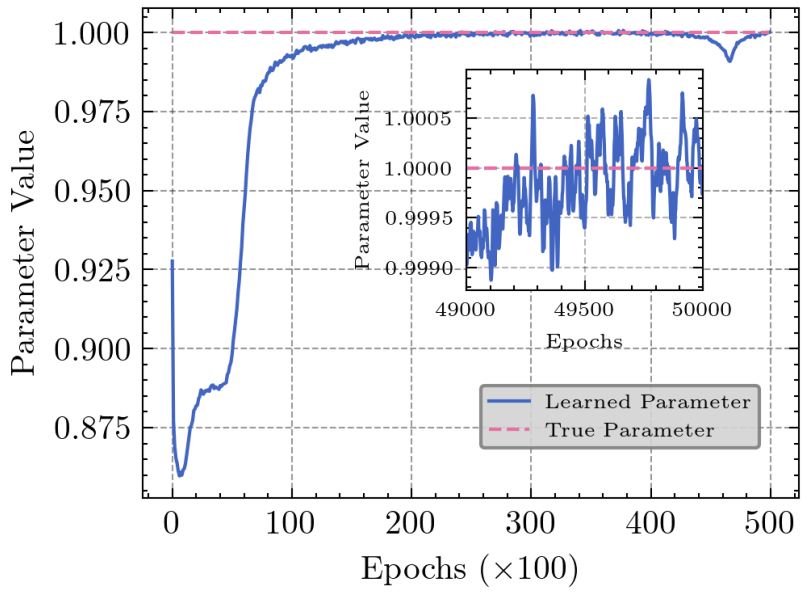


Figure 6.15: Coupling parameter learning, g_1, in the Jaynes-Cummings data simulation and Jaynes-Cummings loss ODE experiment, with initial condition is $|g\rangle \otimes |\alpha\rangle$, where $|\alpha\rangle = e^{-|\alpha|/2} \sum_{k=0}^{10} \frac{\alpha^k}{\sqrt{n!}} |k\rangle$, $\alpha = 1$. The learned coupling converges to a value slightly shifted from the ground truth and stabilizes. The inset shows the convergence of the parameter in the last 1000 epochs of train.

6.2 Effective-Hamiltonian identification

In this set of experiments, we are aiming to identify effective dynamics on a full Hamiltonian. The ODE loss term will use the Rabi Hamiltonian in the interaction picture (Eq. (3.29)), and now the PINN has the task of learning two parameters: g_1 associated with slower oscillations (Jaynes-Cummings remaining term), and g_2 associated with fast oscillations (Rabi exclusive term). In addition to this, a brief discussion on the positivity reinforcement function is made, namely, the absolute value function and the softplus function. All the parameters used in this set of experiments, including data simulation, neural network settings, and training process, are available at Table 6.2.

Table 6.2: Parameters of the simulation, neural network, and training for the experiments simulating data with the Jaynes-Cummings/Rabi Hamiltonians and using the Rabi Hamiltonian in the L_{eq} term.

		Jaynes-Cummings data simulation	Rabi data simulation
Simulation	Atom initial state	$\|e\rangle \otimes \|0\rangle$	$\|e\rangle \otimes \|0\rangle$
	Coupling parameter	$g_1 = 1, g_2 = 0$	$g_1 = g_2 = 1$
	Frequency parameters	$\omega_a = \omega_c = 100$	$\omega_a = \omega_c = 100$
	Hamiltonian frame	interaction	interaction
	Atom dimension	2	2
	Cavity dimension	2	2
	Simulation duration	2π	2π
	Number of data samples	500	500
Network	Neurons in hidden layers	[100, 100, 100]	[100, 100, 100]
	Activation function	sin	sin
	Dropout probability	0	0
	Number of extra parameters	2	2
Training	Number of epochs	50k	50k
	Number of points in L_{data}	1000	1000
	Optimizer	Adam	Adam
	Loss	MAE	MAE
	Learning rate	0.01	0.01
	Bias on coupling	absolute value/softplus	absolute value/softplus

6.2.1 Simulating train data using the Rabi Hamiltonian

In this scenario, train data is simulated using the Rabi Hamiltonian with $\omega_{a,c} \gg g_{1,2}$. We want to check if the neural network is capable to learn that fast-oscillating terms are negligible to the dynamics observed and naturally learn that if $g_2 \to 0$, to a good approximation no dynamics change.

Fig. 6.16 compares the learned and simulated complex amplitudes (real and imaginary parts) of the interaction-frame state, $|\tilde{\psi}(t)\rangle$. The network reproduces the qualitative structure of the wavefunction components over the full time interval, and the residual maps (bottom panels) indicate small absolute discrepancies at the level of 10^{-2} or below across the basis components. This confirms that removing the fast carrier phases substantially alleviates the stiffness of the learning problem, allowing a low-frequency representation to capture the dominant state evolution. Also, it shows that including another parameter in the task of the neural network does not affect its state

learning capability, given the quality of state reconstruction.

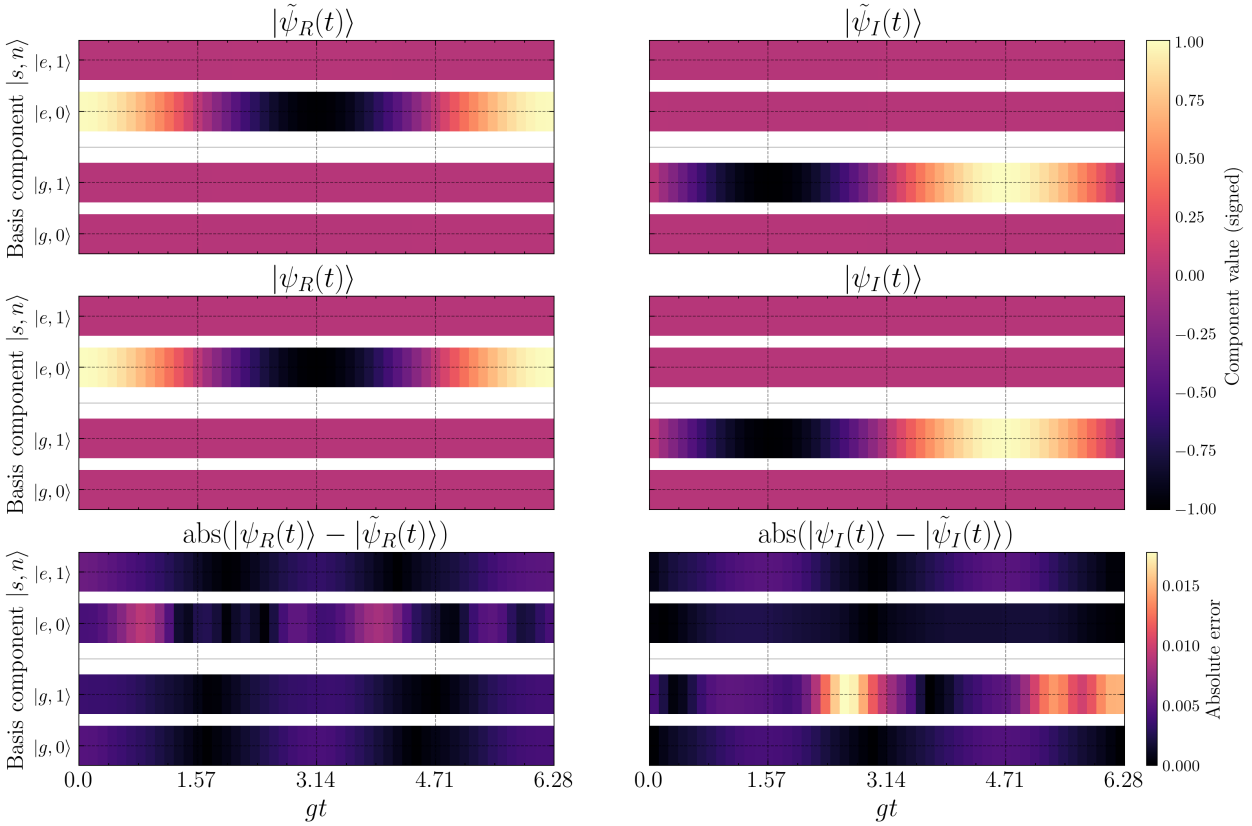


Figure 6.16: State reconstruction in the Rabi data simulation and the Rabi loss ODE experiment, initial condition is $|e\rangle \otimes |0\rangle$. The y-axis shows the basis vectors $|s,n\rangle$ of the composed system (s is associated with the atom and n with the cavity), while the x-axis shows the time evolution in the characteristic gt time scale. Top rows: learned real/imaginary amplitudes. Middle rows: simulated (target) amplitudes. Bottom rows: absolute difference, element-wise. The color bars represent the number value shown in the main plots. The network captures the global dynamical structure, with small errors throughout the simulation time. It is worth noting that the error scale in the real part of the network is around one order of magnitude lower than the error scale in the imaginary part.

Fig. 6.17 shows the time evolution of the key observables used in the data term, namely the cavity photon number $\langle a^\dagger a\rangle$ and the atomic excitation $\langle \sigma_+\sigma_-\rangle$. The predicted curves track the simulated trajectories closely across the entire window, with errors remaining at the percent level or better. Importantly, this agreement is achieved without requiring the network to fit the rapid phase oscillations in the raw Schrödinger-picture amplitudes, reinforcing that those high-frequency features are largely irrelevant for the chosen observables in the regime $\omega_{a,c} \gg g_{1,2}$.

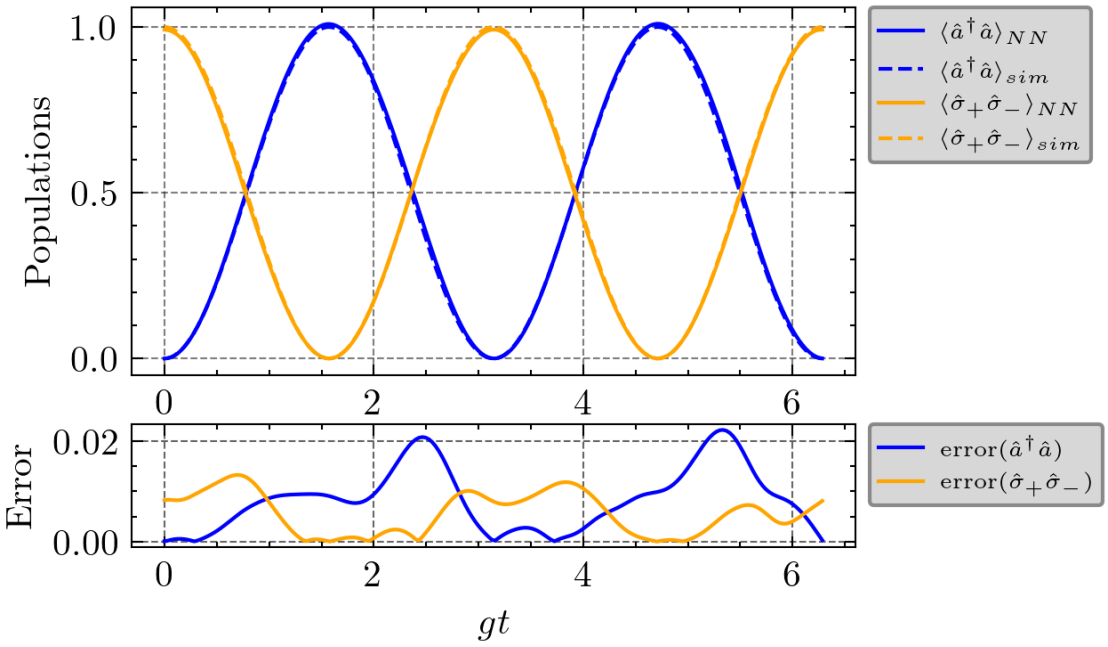


Figure 6.17: Expected values of $\hat{a}^\dagger \hat{a}$ (regarding the cavity) and $\hat{\sigma}_+ \hat{\sigma}_-$ (regarding the atom) operators in the Rabi data simulation and the Rabi loss ODE experiment, with initial condition $|e\rangle \otimes |0\rangle$. Solid lines: PINN predictions computed from $|\tilde{\psi}(t)\rangle$. Dashed lines: simulated targets. The bottom panel reports the signed error $\langle \tilde{\psi}(t)|\hat{O}|\tilde{\psi}(t)\rangle - \langle \psi(t)|\hat{O}|\psi(t)\rangle$ for each observable, remaining small over the full interval.

Fig. 6.18 reports the fidelity between the learned and simulated interaction-frame states. The fidelity remains close to unity throughout the evolution, with a modest dip at $gt \sim 4$, which is consistent with the small but nonzero state-component errors observed in Fig. 6.16, specially for the imaginary part $|\psi_I(t)\rangle$.

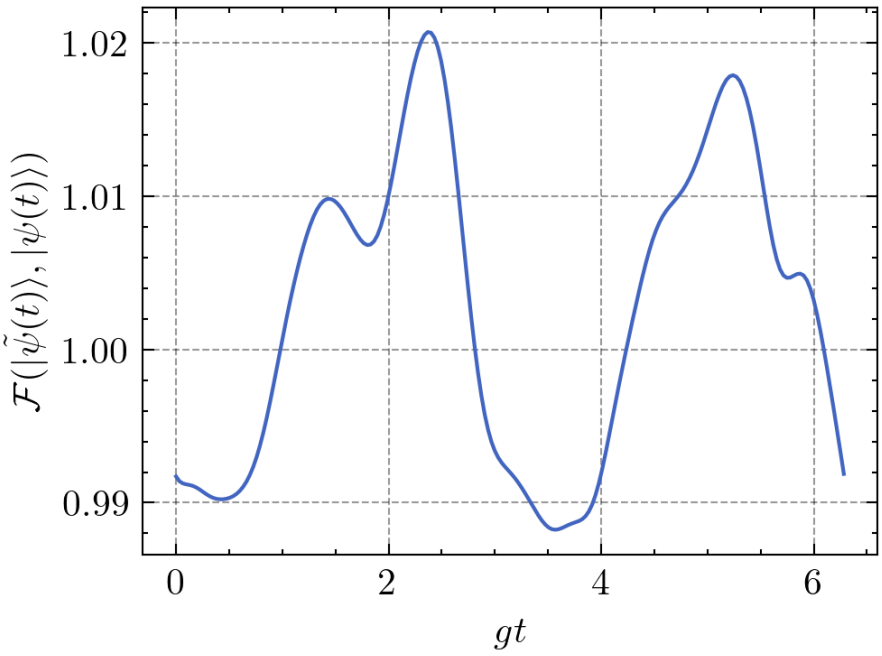


Figure 6.18: Fidelity of the learned state in the Rabi data simulation and the Rabi loss ODE experiment, with initial condition $|g\rangle \otimes |1\rangle$. The values remain close to unity across time, dropping to 0.98 at the end of the simulation, but overall indicating good agreement between simulated and learned states. At some time intervals, we can see that $\mathcal{F} > 1$, which happens because normalization is not a hard constraint in the neural network and small deviations where $\left|\langle\tilde{\psi}(t)|\tilde{\psi}(t)\rangle\right|^2 > 1$ can happen.

Fig. 6.19 displays the evolution of the composite loss and its components. After the initial transient, the data loss term rapidly decreases by several orders of magnitude and then stabilizes alongside the norm and initial conditions losses. They steadily maintain themselves in a low floor throughout the training process.

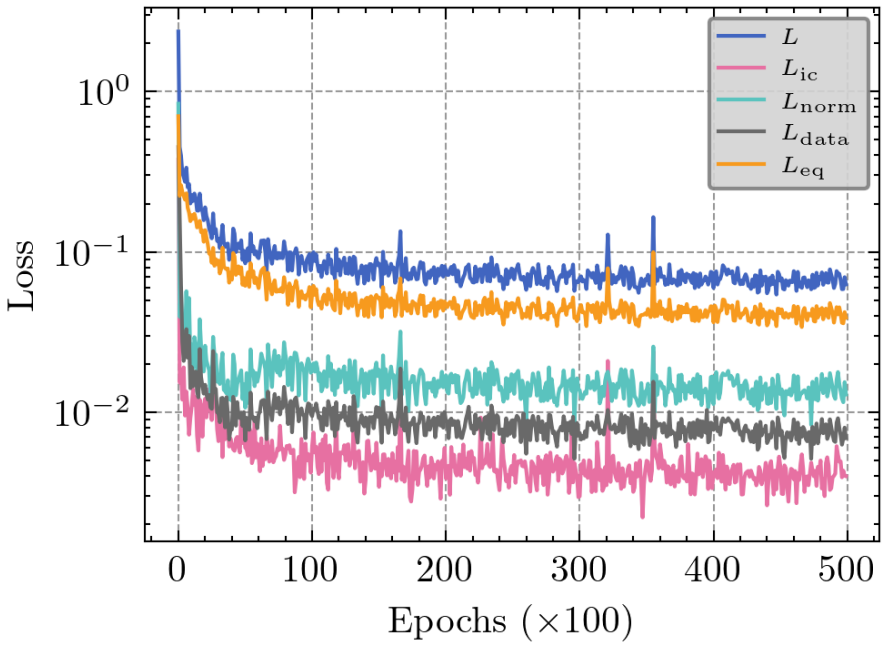


Figure 6.19: Total loss and its components discriminated during the training phase of the Rabi data simulation and the Rabi loss ODE experiment, initial condition is $|e\rangle \otimes |0\rangle$. The total loss and its components decrease several orders of magnitude during training and stabilize, indicating a consistent compromise between data fitting and physics enforcement. All components of the total loss have the same unitary weight.

Fig. 6.20 shows the two experiments of parameter estimation using different inductive bias functions. The network successfully recovers $g_1 \approx 1$ while driving the counter-rotating coupling to $g_2 \approx 0$. This outcome is consistent with the effective Jaynes-Cummings description in the rotating-wave regime: for $\omega_{a,c} \gg g_{1,2}$, the counter-rotating contributions oscillate rapidly in the interaction picture and produce only small corrections to the slow dynamics, so a model with $g_2 \to 0$ can fit the observable trajectories essentially as well as the full model over the investigated time window. In this sense, the training procedure selects an *effective* Hamiltonian representation that is sufficient to reproduce the measured dynamics.

Fig. 6.20(a) shows learned coupling parameters during training with the non-negativity inductive bias $g = |\theta|$ (absolute value). The Jaynes-Cummings coupling g_1 rapidly converges to the target value while the counter-rotating coupling g_2 is driven to (and remains at) values numerically indistinguishable from zero, consistent with an effective Jaynes-Cummings description in the rotating-wave regime. The inset highlights the final epochs, showing stable convergence.

Fig. 6.20(b) shows learned coupling parameters during training with the smooth non-negativity map $g = \text{softplus}(\theta)$. Both parameters remain strictly nonnegative and evolve smoothly; g_1 converges to the target value while g_2 decreases but typically approaches a small positive floor rather than collapsing exactly to zero, reflecting the asymptotic behavior of the softplus parameterization.

The inset emphasizes the late-training plateau.

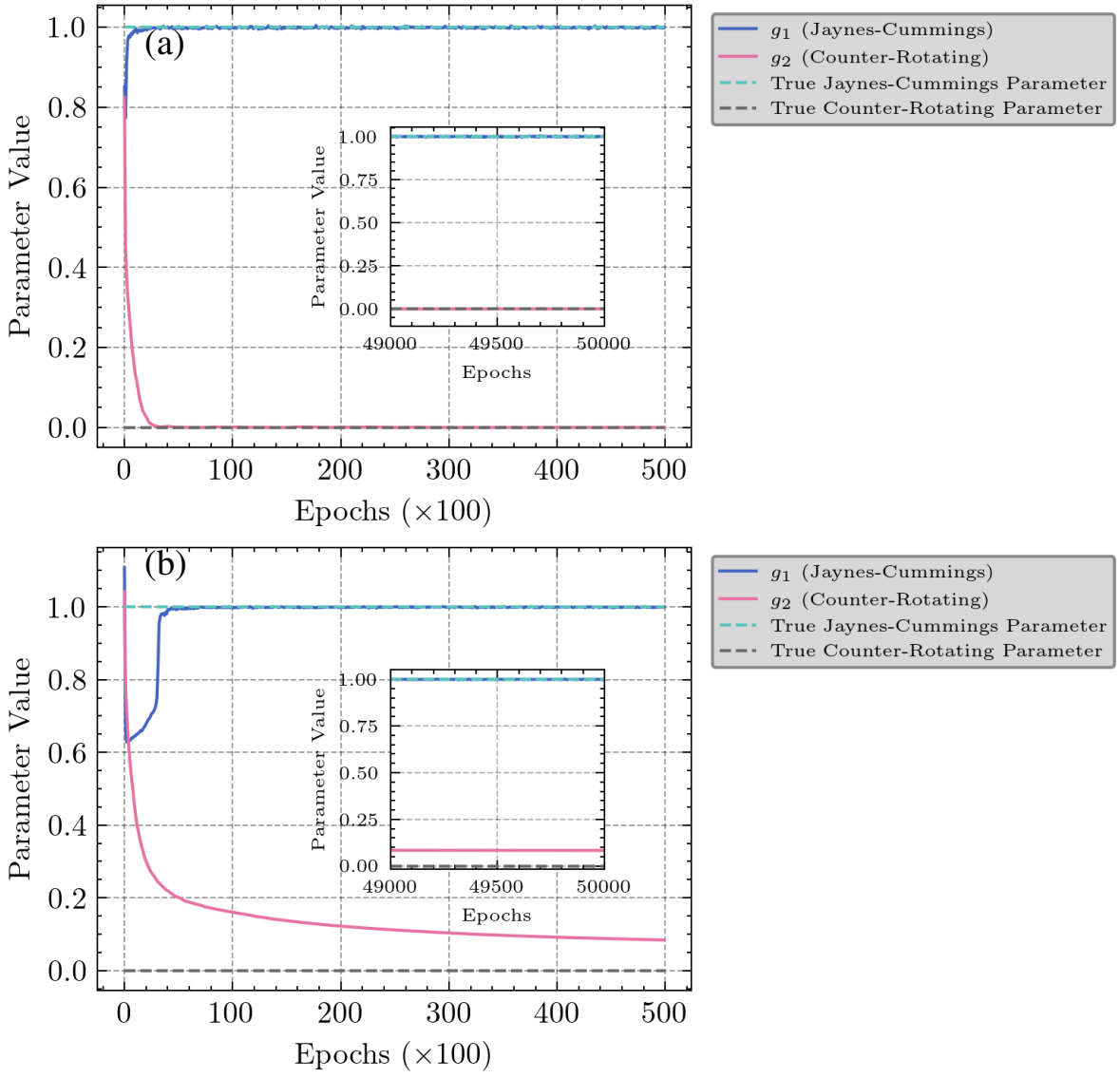


Figure 6.20: Training trajectories of the learned interaction couplings for the split Rabi Hamiltonian, comparing two non-negativity parameterizations in the Rabi data simulation and Rabi ODE loss term experiment, with initial condition $|e\rangle \otimes |0\rangle$. (a) uses the absolute-value map $g = |\theta|$, which permits boundary solutions and can promote pruning of weakly identifiable terms (here $g_2 \to 0$). (b) uses the smooth softplus map $g = \text{softplus}(\theta)$, which yields smoother optimization but enforces strictly positive couplings unless driven to large negative pre-activations. In both cases, g_1 converges to the Jaynes-Cummings coupling value, while the evolution of g_2 reveals how the chosen inductive bias affects the emergence of an effective Jaynes-Cummings model.

Although both choices on non-negativity of the parameter guarantee $g \geq 0$, they differ in optimization behavior and in how easily a parameter can be driven toward (or away from) zero. The absolute value map is *non-differentiable at* $\theta = 0$ and has gradient $|\theta|' = \text{sign}(\theta)$ for $\theta \neq 0$. In

practice, this yields a piecewise-constant gradient magnitude with a kink at the origin, as can be seen in Fig. 6.21(a). Consequently, it is comparatively easy for gradient descent to move g_2 toward 0 and, once sufficiently close, numerical updates may cause it to *stick* near 0 (especially if updates alternate sign around the kink). This promotes a phenomenon in machine learning called *pruning*, in which the parameters of the model being trained are set to zero because they contribute little to the predictions made [67, 68].

In contrast, softplus(θ) is smooth and strictly positive, with derivative softplus$'(\theta) = 1/(1+e^{-\theta})$. By construction, this implies smoother gradients and typically more stable optimization; and g approaches zero only asymptotically as $\theta \to -\infty$, so the learned g_2 will often become *very small* rather than exactly zero. Fig. 6.21(b) shows the softplus curve and its derivative.

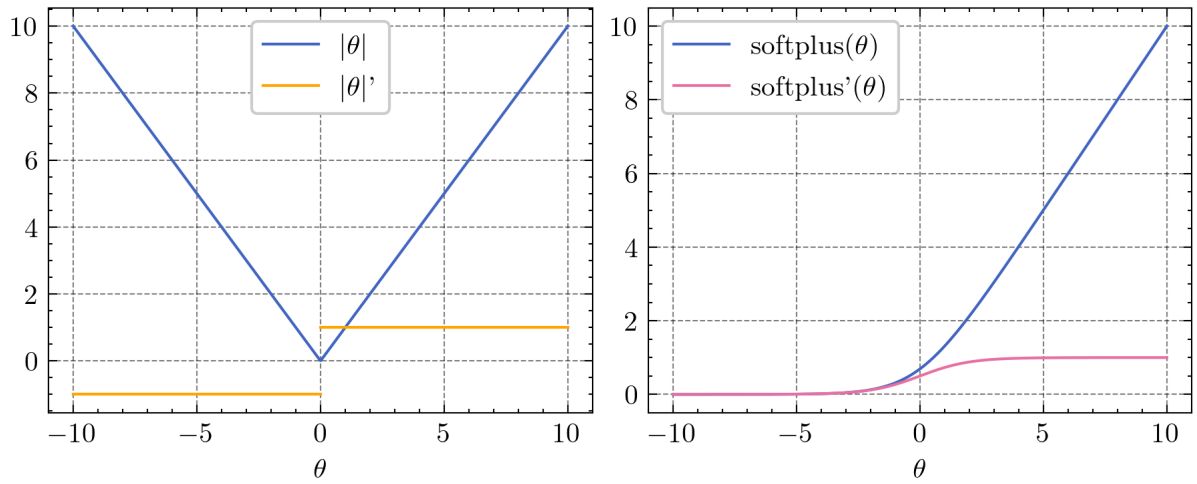


Figure 6.21: Two possible inductive bias applied to the learned parameter. Both are used to enforce a positive parameter, but they imply in different loss landscapes. (a) Absolute value function and its derivative. (b) Softplus function and its derivative. In the Rabi model context, the absolute value function shown on (a) strongly enforces that $g_2 \to 0$, while the softplus function on (b) is less effective at parameter pruning.

Therefore, the fact that g_2 collapses to (nearly) zero in Fig. 6.20 is consistent with two combined effects: (i) in the rotating-wave regime the counter-rotating term is weakly identifiable from the chosen observables and time horizon, producing a relatively flat loss direction in g_2; and (ii) the absolute-value parameterization introduces a kink at zero that can facilitate convergence to a boundary solution. Therefore, using $g = |\theta|$ can *accentuate* the tendency to select an effective Jaynes-Cummings model by making the boundary $g_2 = 0$ numerically attractive.

6.2.2 Simulating train data using the Jaynes-Cummings Hamiltonian

In this scenario, train data are simulated using the Jaynes-Cummings Hamiltonian with $\omega_{a,c} \gg g_1$. We want to check whether the neural network is able to learn that the data provided are in agreement with an effective dynamics where the fast-oscillating terms associated with the coupling term g_2 are negligible.

Fig. 6.22 compares the learned interaction-frame wavefunction components with the simulated Jaynes-Cummings targets. The network accurately reproduces the dominant structure of the real and imaginary parts of $|\tilde{\psi}(t)\rangle$ over the full time window, and the absolute error maps (bottom panels) indicate uniformly small discrepancies, remaining at the 10^{-3}–10^{-2} level across the basis components. This agreement shows that, despite being trained with the more expressive Rabi hypothesis class, the PINN converges to a state evolution consistent with the Jaynes-Cummings-generated data.

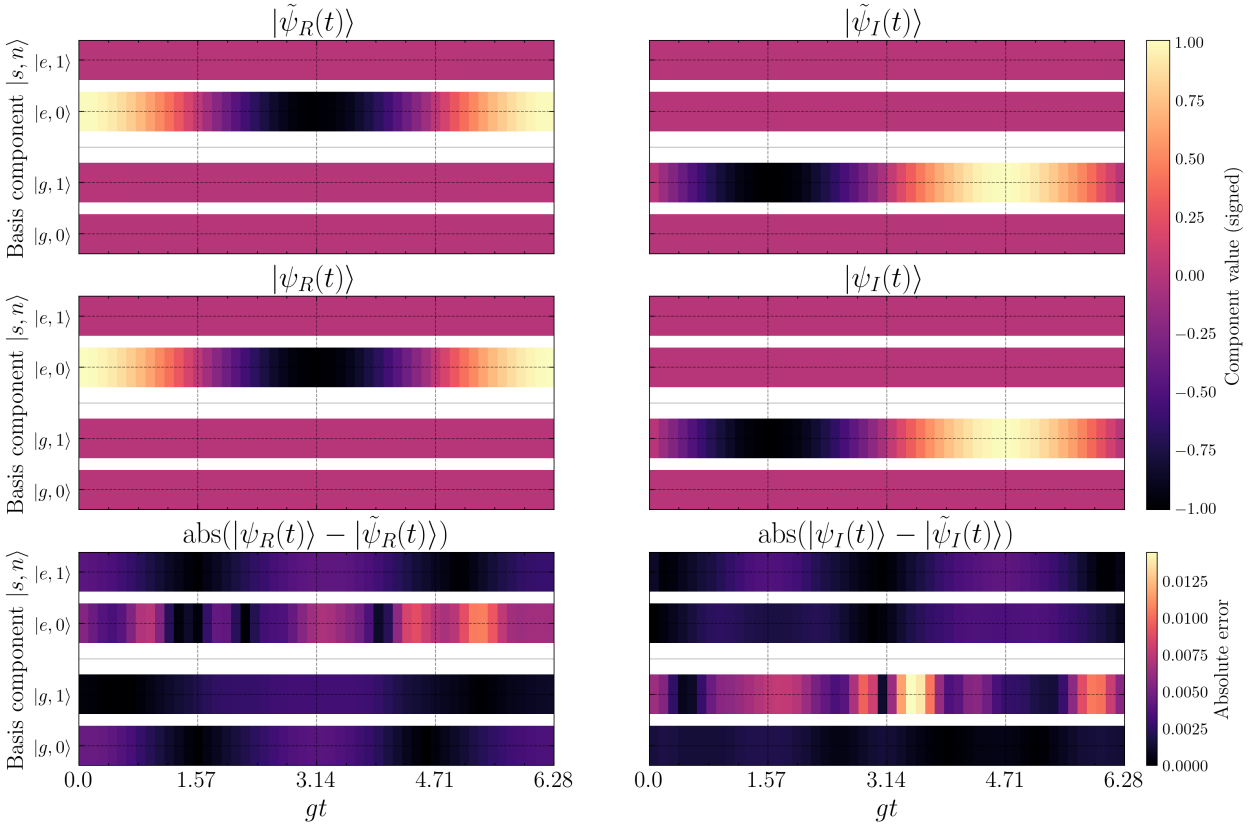


Figure 6.22: State reconstruction in the Jaynes-Cummings data simulation and the Rabi loss ODE experiment, with initial condition $|e\rangle \otimes |0\rangle$. Each plot horizontally shows the time evolution, and vertically shows the elements of the state vector; each strip is one of the vector elements evolving in time. Top rows: learned real/imaginary amplitudes. Middle rows: simulated (target) amplitudes. Bottom rows: absolute difference, element-wise. The network captures the global dynamical structure, with small errors throughout the simulation time.

The predicted expectation values for the photon number and atomic excitation closely track the simulated curves, as shown in Fig. 6.23. The corresponding error traces remain small and bounded, indicating that the learned state is not only visually consistent with the target amplitudes but also reproduces the physically relevant observables used in the data loss. Complementarily, the fidelity between the learned and simulated states (Fig. 6.24) remains high throughout the evolution, with only modest fluctuations around values close to unity, displaying a slight drop near $gt \approx 5$, when the imaginary part of the learned state vector most disagrees with the ground truth, upon close inspection of Fig. 6.22.

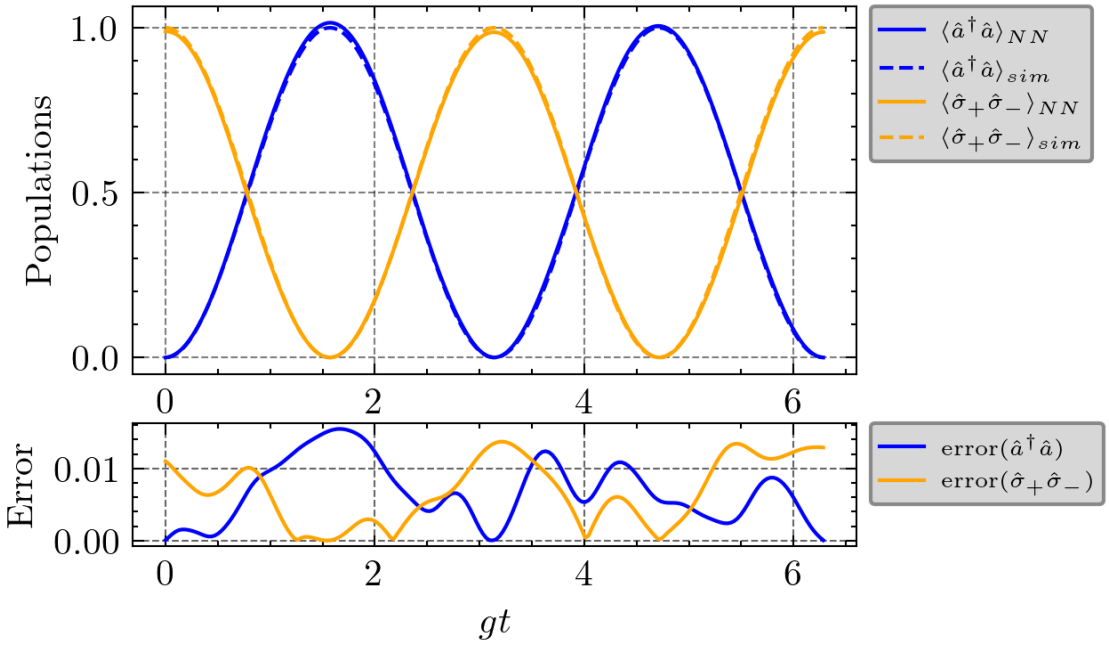


Figure 6.23: Expected values of $\hat{a}^\dagger \hat{a}$ (regarding the cavity) and $\hat{\sigma}_+ \hat{\sigma}_-$ (regarding the atom) operators in the Jaynes-Cummings data simulation and the Rabi loss ODE experiment, with initial condition $|e\rangle \otimes |0\rangle$. Solid lines: PINN predictions computed from $|\tilde{\psi}(t)\rangle$. Dashed lines: simulated targets. The bottom panel reports the signed error $\left| \langle \tilde{\psi}(t)|\hat{O}|\tilde{\psi}(t)\rangle - \langle \psi(t)|\hat{O}|\psi(t)\rangle \right|$ for each observable, remaining small over the full interval.

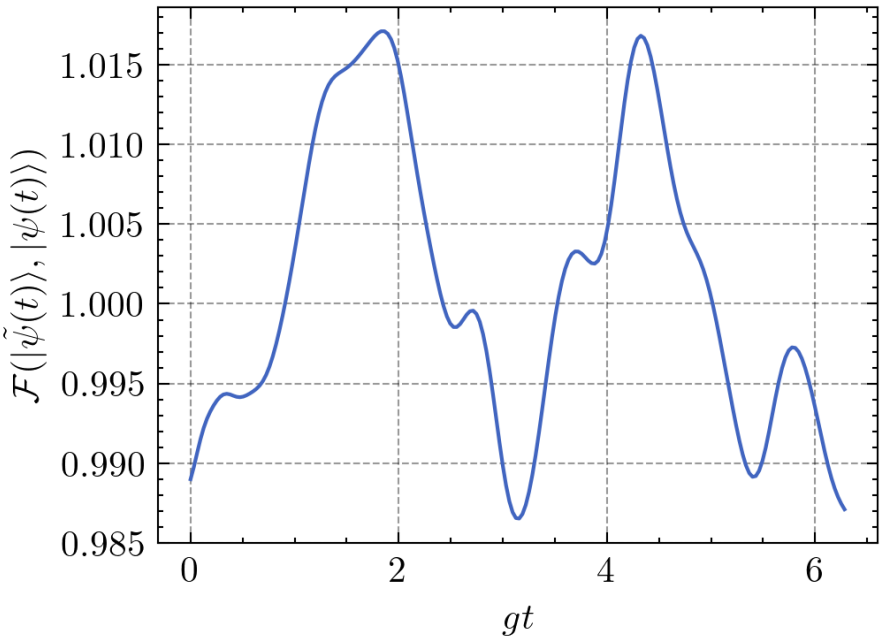


Figure 6.24: Fidelity of the learned state in the Jaynes-Cummings data simulation and the Rabi loss ODE experiment, with initial condition $|e\rangle \otimes |0\rangle$. The values remain close to unity across time, dropping to 0.985 in the first half of the simulation, but overall indicating good agreement between simulated and learned states. At some time intervals, we can see that $\mathcal{F} > 1$, which happens because normalization is not a hard constraint in the neural network and small deviations where $\left| \langle \tilde{\psi}(t) | \psi(t) \rangle \right|^2 > 1$ can happen.

Fig. 6.25 shows a rapid reduction of the ODE residual followed by a plateau where the remaining loss is dominated by a balance among the initial-condition, normalization, and data terms. A drop in the total loss is observed around 10k epochs into the training loop mainly driven by a drop in the L_{data} term, indicating a transition to a lower-error basin where the network satisfies the observational supervision more effectively. This figure displays a behavior encountered in many experiments, which is the antagonism played by L_{eq} and L_{data} terms: when the L_{data} term suddenly drops, it is at the expense of the L_{eq} peaking, probably due to the neural network struggling to find balance between these two terms. We observe that, as the training process evolves, both terms reach a stable regime in a low-valued floor.

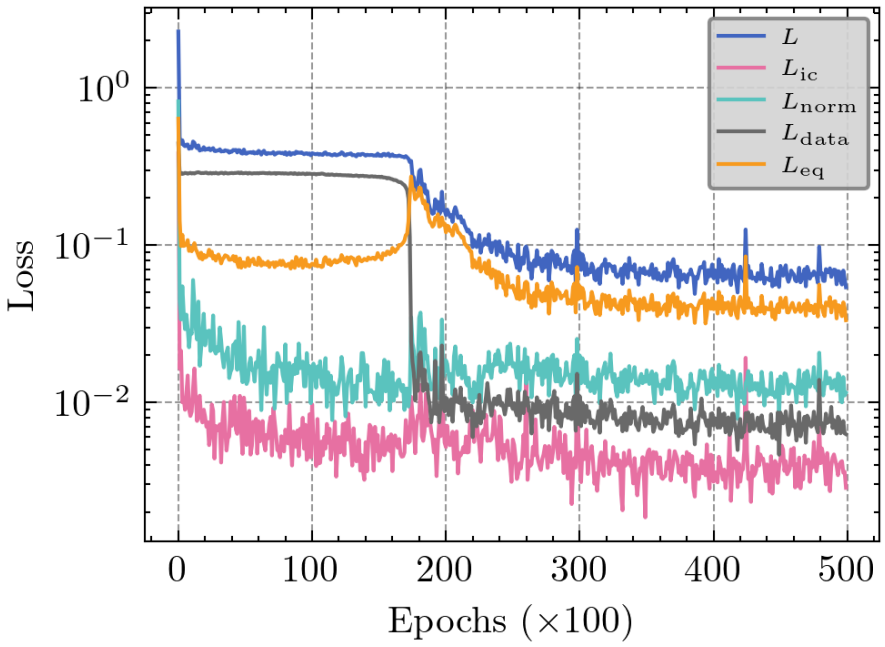


Figure 6.25: Total loss and its components discriminated during the training phase of the Jaynes-Cummings data simulation and the Rabi loss ODE experiment. Initial condition is $|e\rangle \otimes |0\rangle$. The L_{eq} and L_{data} terms display an interesting dynamics where the former peaks at around epoch 10k, while the later drops almost at the same time. That indicates a tendency of the neural network to prioritize data matching over dynamics matching, which eventually balances during the training process. All terms in the loss function have the same unitary weight.

Fig. 6.26(a) provide the key diagnostic of model selection. Since the data are generated with $g_2 = 0$, any nonzero g_2 in Eq. (3.29) would represent an unnecessary interaction term. The PINN indeed drives the counter-rotating parameter toward values close to zero, while g_1 converges to the expected Jaynes-Cummings coupling (as highlighted by the inset). This behavior is consistent with the physical picture of the rotating-wave approximation: in the interaction frame, the counter-rotating contribution oscillates at $\omega_c + \omega_a$ and tends to average out when $\omega_{a,c} \gg g_{1,2}$, so that an effective Hamiltonian with $g_2 \simeq 0$ can reproduce the measured dynamics.

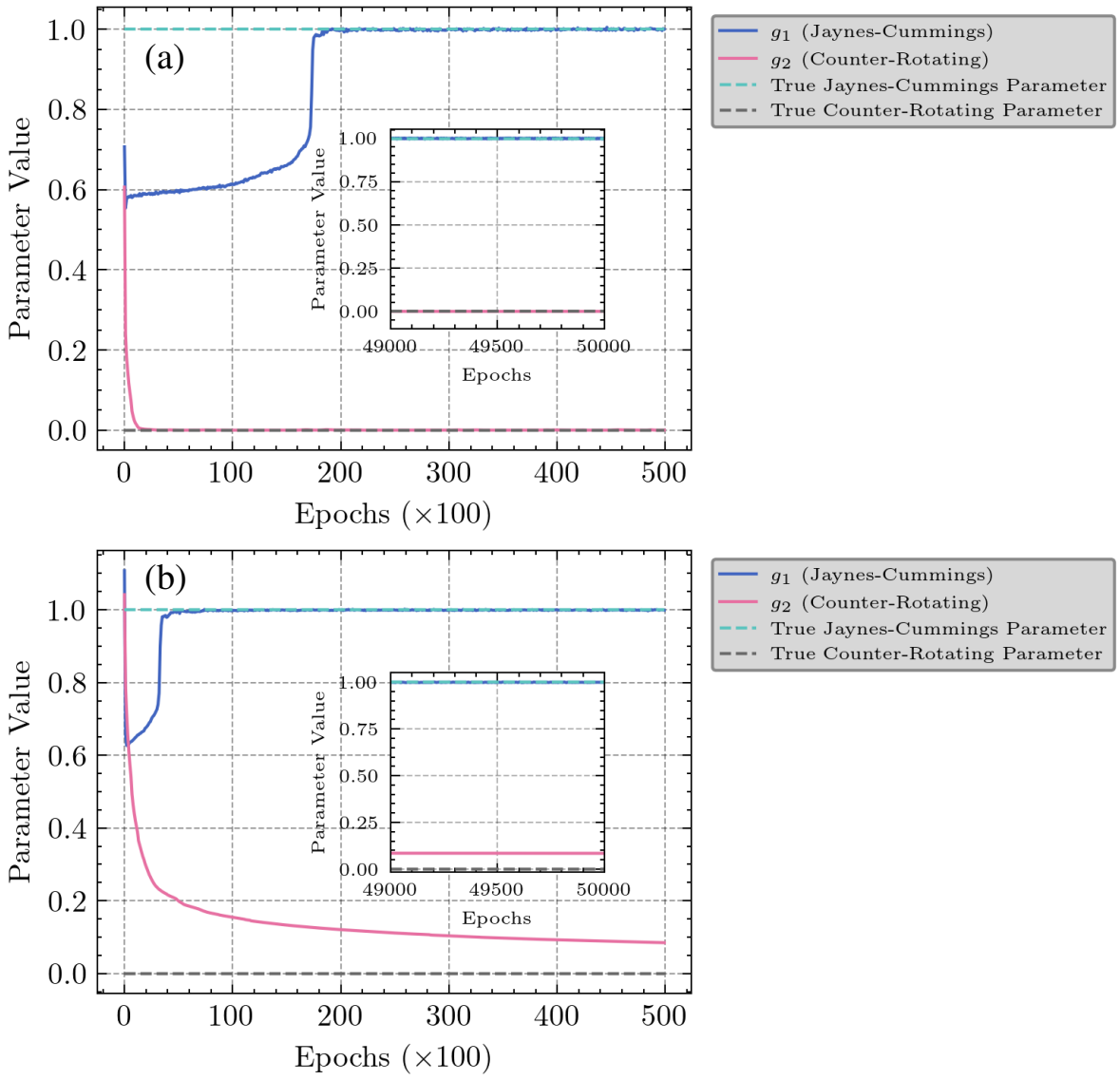


Figure 6.26: Training trajectories of the learned interaction couplings, comparing two non-negativity parameterizations in the Jaynes-Cummings data simulation and Rabi ODE loss term experiment, with initial condition $|e\rangle \otimes |0\rangle$. (a) uses the absolute-value map $g = |\theta|$, which permits boundary solutions and can promote pruning of weakly identifiable terms (here $g_2 \to 0$). (b) uses the smooth softplus map $g = \text{softplus}(\theta)$, which yields smoother optimization but enforces strictly positive couplings. Similar to Fig. 6.20, g_1 converges to the Jaynes-Cummings coupling value, while the discrepancy in the evolution of g_2 shows the difference between inductive bias function chosen.

Interestingly, the parameter trace for g_1 exhibits a transient regime where it remains at lower values before abruptly rising and locking near the correct value. This "delayed identification" is consistent with an optimization landscape where early training prioritizes satisfying the ODE constraint and normalization with a low-amplitude effective interaction, followed by a later stage

in which the data term forces the model into the correct coupling basin. In other words, the PINN first learns a dynamically self-consistent but weakly interacting trajectory, and only later adjusts the coupling strength to match the supervised observables. This effect is also in accordance with Fig. 6.25, although none of these behaviors were explicitly enforced in the training process, and other training iterations showed different outcomes in the learned parameters trajectory.

In the other hand, Fig. 6.26(b) shows the parameter learning curve using $g_2 = \text{softplus}(\theta_2)$. In this run, g_1 converges much more rapidly than what is shown in Fig. 6.26(a), and we can clearly identify an almost monotonic trend of $g_2 \to 0$.

Overall, this experiment demonstrates that training with the more general Rabi Hamiltonian loss ODE term using Jaynes-Cummings-simulated data does not necessarily corrupt the learned dynamics; rather, the PINN can converge toward an *effective* Jaynes-Cummings description by suppressing the counter-rotating coupling. From a model-discovery perspective, this constitutes evidence that the framework can recover a simpler effective interaction from within a larger hypothesis class when the data are consistent with the rotating-wave regime. At the same time, the transient behavior in the parameter trajectory highlights that the identification of weakly constrained parameters may depend sensitively on the loss weighting and the optimization path, motivating the use of systematic schedules (e.g. curriculum strategies or adaptive weighting [69]) when extending to regimes where counter-rotating effects become significant.

Chapter 7

Conclusion

This dissertation investigated whether PINNs can be used as a practical tool for effective Hamiltonian learning in quantum dynamics. Motivated by the central role of effective descriptions in quantum optics and cavity QED, the Rabi and Jaynes-Cummings models were adopted as a proof-of-concept setting in which the passage from a full microscopic model to an effective one is both physically meaningful and theoretically well established.

The framework developed here combines a neural surrogate for the quantum state with physics constraints derived from Schrödinger dynamics. Working in the interaction picture, the complex-valued state is represented by two real-valued neural networks (real and imaginary components), trained jointly with unknown coupling parameters. The training objective couples a data term built from experimentally accessible expectation values to an ODE residual evaluated at different time instants, together with initial-condition and normalization enforcement.

In the matched-model setting (Jaynes-Cummings simulated data and Jaynes-Cummings ODE residual), the inverse-PINN formulation was used to jointly reconstruct the state evolution and infer the coupling strength using only a limited set of expectation-value from observables. Importantly, robustness to the initial cavity state was explicitly examined by testing distinct families of initial conditions (including Fock-type and coherent states), providing a structured evaluation of how the parameter-learning task depends on the dynamical richness induced by the initial state.

A key conceptual objective of this dissertation was to move beyond parameter fitting within a fixed model, and instead investigate whether physics-informed learning can help bridge full and effective Hamiltonian descriptions. To that end, the split Rabi interaction Hamiltonian was used in the ODE residual with two learnable couplings that separately weight rotating and counter-rotating processes. In the weak-coupling, resonant regime considered, the reported results show that training can suppress the counter-rotating small contribution and converge toward an effective Jaynes-Cummings description, thereby supporting the idea that inverse PINNs can be used to test whether terms in dynamics generators are warranted by the data.

Finally, in the mismatched-model experiment where data are generated by Jaynes-Cummings dynamics while the PINN is trained with the more general split Rabi model, the reported parameter trajectories indicate that the counter-rotating coupling is driven close to zero while the rotating coupling converges close to its target value. From the standpoint of model reduction, this behavior is significant: even when trained with a more expressive Hamiltonian class, the inverse-PINN procedure can "switch off" unsupported terms and recover an effective generator consistent with the data.

A practical finding of this study is that inductive bias in the parameterization can influence the outcome of parameter discovery. When enforcing positivity of the learned couplings, the absolute-value transformation and the softplus transformation can lead to different convergence behaviors, with the absolute-value parameterization reported to converge to values closer to the effective-model expectation in the considered setting. This emphasizes that, in inverse problems, the learning outcome reflects both the physical constraints and the design choices made in encoding them.

7.1 Limitations

While the results presented here provide a clear proof of concept, there are important limitations that delimit the scope of the present conclusions.

- The experiments rely on simulated expectation values under controlled conditions. Real experiments introduce noise, drift, calibration effects, and imperfect initialization, all of which may affect identifiability and training stability.

- Due to the nature of this study (a proof of concept employing a well-documented and tested system), strong previous knowledge was used in order to find the right convergences. For instance, dividing the Rabi Hamiltonian into two terms (one with slow oscillations and the other one with fast oscillations) already points out to the finish line, which is the Jaynes-Cummings Hamiltonian.

- The study focuses on resonant dynamics in selected coupling regimes and employs a finite-dimensional truncation of the cavity Hilbert space. Extensions to detuned, stronger-coupling, or higher-excitation regimes may require additional modeling and computational care.

7.2 Future directions

The framework developed in this dissertation opens several clear and impactful directions for future work.

- A natural next step is to incorporate realistic noise models and evaluate robustness under sparser sampling, missing time windows, and imperfect initial conditions.
- Many realistic cavity-QED platforms are dissipative. Extending the physics constraint from unitary Schrödinger evolution to Lindblad master-equation dynamics would enable simultaneous inference of coherent parameters (couplings, detunings) and incoherent rates (decay, dephasing), and would directly connect with the broader motivation for effective descriptions.
- Relaxing the priors about the system are a necessary next step to reach a more general framework, such as assuming no separability of the full Hamiltonian into terms that carry this or that behavior with the system.
- To explore the technique developed, there is room to extend this frameworks to other settings in the same atom-cavity context, for instance: obtain the effective Hamiltonian in non-resonant regimes, and explore higher dimensional Hilbert spaces in the atom subsystem.
- To investigate adaptive re-weighting of the loss function terms. Training neural networks to optimize many loss terms at the same time can be challenging, specially when they antagonize each other at times, so learning efficient gradient-based ways of dynamically giving more importance to certain terms than others during training could lead to better neural network predictions [70].

7.3 Final Remarks

In conclusion, this dissertation establishes a coherent physics-informed learning workflow for Hamiltonian parameter inference in quantum light-matter systems using only partial observation channels, and it demonstrates how the same framework can be used to interrogate the consistency of additional Hamiltonian terms with measured dynamics. By framing effective-Hamiltonian discovery as an inverse problem constrained by physical laws, the approach developed here offers a promising pathway toward data-driven model validation and, ultimately, model discovery in quantum optics and related experimental platforms.

Bibliography

[1] T. H. Maiman, Nature **187**, 493 (1960).

[2] C. H. Bennett and G. Brassard, Theoretical Computer Science **560**, 7 (2014), Archival journal publication of the original 1984 manuscript (Proc. IEEE Int. Conf. Computers, Systems and Signal Processing, Bangalore, Dec. 1984).

[3] J. Aasi *et al.*, Nature Photonics **7**, 613 (2013).

[4] J. Stuhler and A. J. Shields, Problems and Solutions on Optics (2019).

[5] C. Fabre and N. Treps, Rev. Mod. Phys. **92**, 35005 (2020).

[6] D. F. Walls and G. J. Milburn, *Quantum Optics* (Springer, Berlin, Heidelberg, 1994).

[7] C. C. Gerry and P. L. Knight, *Introductory Quantum Optics* (Cambridge University Press, Cambridge, 2004).

[8] I. I. Rabi, Physical Review **51**, 652 (1937).

[9] A. Pereverzev and E. R. Bittner, Physical Chemistry Chemical Physics **8**, 1378 (2006).

[10] D. Braak, Physical Review Letters **107**, 100401 (2011).

[11] Q. Xie, H. Zhong, M. T. Batchelor, and C. Lee, Journal of Physics A: Mathematical and Theoretical **50**, 113001 (2017).

[12] M. O. Scully and M. S. Zubairy, *Quantum optics* (Cambridge university press, 1997).

[13] E. T. Jaynes and F. W. Cummings, Proceedings of the IEEE **51**, 89 (1963).

[14] S. Haroche and J.-M. Raimond, *Exploring the Quantum: Atoms, Cavities, and Photons* (Oxford University Press, Oxford, 2006).

[15] A. Blais, R.-S. Huang, A. Wallraff, S. M. Girvin, and R. J. Schoelkopf, Physical Review A **69**, 062320 (2004).

[16] J. Larson and T. Mavrogordatos, *The Jaynes–Cummings Model and its Descendants (Second Edition)*2053-2563 (IOP Publishing, 2024).

[17] C. Villas-Boas and D. Z. Rossatto, Physical review letters **122 12**, 123604 (2018).

[18] P. Forn-Díaz, L. Lamata, E. Rico, J. Kono, and E. Solano, Reviews of Modern Physics **91**, 025005 (2019).

[19] A. B. Klimov, L. L. Sánchez-Soto, A. Navarro, and E. C. Yustas, Journal of Modern Optics **49**, 2211 (2002).

[20] D. F. V. James and J. Jerke, Canadian Journal of Physics **85**, 625 (2007).

[21] M. I. Jordan and T. Mitchell, Science **349**, 255 (2015).

[22] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning* (MIT press, 2016).

[23] A. Halevy, P. Norvig, and F. Pereira, IEEE Intelligent Systems **24**, 8 (2009).

[24] Y. LeCun, Y. Bengio, and G. Hinton, Nature **521**, 436 (2015).

[25] A. Krizhevsky, I. Sutskever, and G. E. Hinton, Communications of the ACM **60**, 84 (2017).

[26] A. Vaswani *et al.*, Attention is all you need, 2017, 1706.03762.

[27] D. Silver *et al.*, Nature **529**, 484 (2016).

[28] A. Esteva *et al.*, Nature **542**, 115 (2017).

[29] J. Jumper *et al.*, Nature **596**, 583 (2021).

[30] W. S. McCulloch and W. Pitts, The bulletin of mathematical biophysics **5**, 115 (1943).

[31] F. Rosenblatt, Psychological review **65**, 386 (1958).

[32] P. J. Werbos, *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences*, PhD thesis, Harvard University, Cambridge, MA, 1974.

[33] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, Nature **323**, 533 (1986).

[34] S. L. Brunton, Acta Mechanica Sinica , 1 (2022).

[35] R. Ranade, C. Hill, and J. Pathak, Computer Methods in Applied Mechanics and Engineering **378**, 113722 (2021).

[36] D. Kochkov *et al.*, Proceedings of the National Academy of Sciences **118**, e2101784118 (2021).

[37] M. L. Piscopo, M. Spannowsky, and P. Waite, Phys. Rev. D **100**, 016002 (2019).

[38] J. Blechschmidt and O. G. Ernst, GAMM-Mitteilungen **44**, e202100006 (2021).

[39] C. Michoski, M. Milosavljević, T. Oliver, and D. R. Hatch, Neurocomputing **399**, 193 (2020).

[40] M. Raissi, P. Perdikaris, and G. Karniadakis, Journal of Computational Physics **378**, 686 (2019).

[41] Z. K. Lawal, H. Yassin, D. T. C. Lai, and A. C. Idris, Big Data and Cognitive Computing **6**, 140 (2022).

[42] G. E. Karniadakis *et al.*, Nature Reviews Physics **3**, 422 (2021).

[43] S. Cuomo *et al.*, arXiv preprint arXiv:2201.05624 (2022).

[44] E. O. Oluwasakin and A. Q. M. Khaliq, Algorithms **16**, 547 (2023).

[45] J. Liu and X. Wang, arXiv preprint (2025), 2506.10379.

[46] C. W. Gardiner and P. Zoller, *Quantum Noise: A Handbook of Markovian and Non-Markovian Quantum Stochastic Methods with Applications to Quantum Optics*, 3 ed. (Springer, 2004).

[47] A. A. Clerk, M. H. Devoret, S. M. Girvin, F. Marquardt, and R. J. Schoelkopf, Reviews of Modern Physics **82**, 1155 (2010).

[48] H.-P. Breuer and F. Petruccione, *The Theory of Open Quantum Systems* (Oxford University Press, Oxford, 2002).

[49] H. M. Wiseman and G. J. Milburn, *Quantum Measurement and Control* (Cambridge University Press, Cambridge, 2010).

[50] B. W. Shore and P. L. Knight, Journal of Modern Optics **40**, 1195 (1993).

[51] C. Cohen-Tannoudji, B. Diu, and F. Laloë, *Quantum Mechanics, Volume 1* (Wiley, New York, 1991).

[52] R. Juárez-Amaro and H. M. Moya-Cessa, Applied Mathematics & Information Sciences **9**, 299 (2015).

[53] L. Mandel and E. Wolf, *Optical Coherence and Quantum Optics* (Cambridge University Press, 1995).

[54] T. Werlang, A. Dodonov, E. Duzzioni, and C. Villas-Bôas, Physical Review A—Atomic, Molecular, and Optical Physics **78**, 053805 (2008).

[55] J. J. Sakurai, *Modern Quantum Mechanics* (Addison-Wesley, Reading, MA, 1994).

[56] M. A. Nielsen, *Neural Networks and Deep Learning* (Determination Press, San Francisco, CA, USA, 2015).

[57] D. A. Roberts, S. Yaida, and B. Hanin, *The Principles of Deep Learning Theory* (Cambridge University Press, Cambridge, MA, USA, 2022).

[58] T. Nishijima, arXiv preprint arXiv:2102.10993 (2021).

[59] M. Kubat, *An introduction to machine learning* (Springer, 2017).

[60] H. Robbins and S. Monro, The Annals of Mathematical Statistics **22**, 400 (1951).

[61] L. Bottou, Proceedings of COMPSTAT , 177 (2010).

[62] D. P. Kingma and J. Ba, International Conference on Learning Representations (ICLR) (2015).

[63] A. C. Wilson, R. Roelofs, M. Stern, N. Srebro, and B. Recht, Advances in Neural Information Processing Systems (NeurIPS) **30** (2017).

[64] H. Zhao, F. Liu, L. Li, and C. Luo, Applied Intelligence **48**, 1707 (2018).

[65] J. Johansson, P. Nation, and d. F. Nori, Comput. Phys. Commun **184**, 1234 (2013).

[66] R. J. Glauber, Physical Review **131**, 2766 (1963).

[67] R. Tibshirani, Journal of the Royal Statistical Society: Series B (Methodological) **58**, 267 (1996).

[68] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2 ed. (Springer, 2009).

[69] B. Gao, R. Yao, and Y. Li, Comput. Math. Appl. **181**, 216 (2025).

[70] Z. Hao *et al.*, arXiv preprint arXiv:2211.08064 (2022).

[71] J. R. R. Tolkien, *The Fellowship of the Ring*, The Lord of the Rings Vol. 1 (George Allen & Unwin, London, 1954).