Application of machine learning techniques to the description of open quantum systems

Kimara Naicker

A thesis submitted in fulfilment of the requirement for the degree of

Doctor of Philosophy in Physics

School of Chemistry and Physics

College of Agriculture, Engineering and Science

University of KwaZulu-Natal

March, 2023

As the candidate's supervisor, I have approved this thesis for submission.

Signed:████████████..........Date:2023-03-30..

As the candidate's co-supervisor, I have approved this thesis for submission.

Signed:....████████████.......Date:2023-03-30..

Declaration 1 - Plagiarism

I, **Kimara Naicker**, declare that;

1. The research reported in this thesis, except where otherwise indicated, is my original research.

2. This thesis has not been submitted for any degree or examination at any other university.

3. This thesis does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.

4. This thesis does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:

   (a) Their words have been re-written but the general information attributed to them has been referenced,

   (b) Where their exact words have been used, then their writing has been placed in italics and inside quotation marks, and referenced.

5. This thesis does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the thesis and in the references sections.

Signed:█████████.........Date:30 March 2023..

Declaration 2 - Publication

The following journal papers emanating from this work have been either published or are under review:

1. Machine learning for excitation energy transfer dynamics, K. Naicker, I. Sinayskiy and F. Petruccione, Physical Review Research, 14, 033175 (2022)

2. Statistical and machine learning approaches for prediction of long-time excitation energy transfer dynamics, K. Naicker, I. Sinayskiy and F. Petruccione, arXiv:2210.14160v2 (*submitted for review*)

Other journal papers produced during the period of registration for this study:

1. Charge and spin state of dilute Fe in NaCl and LiF, H. P. Gunnlaugsson, *et al.*, Physical Review B (*accepted*)

2. Temperature-dependence of the magnetic field at Fe sites in Ba-doped BiFeO 3 thin films studied by Emission Mössbauer Spectroscopy, K. Naicker, *et al.* (*in preparation*)

Signed:.... ....

Acknowledgements

Sylvain my fiancé, I thank you for your motivation, your support and for being my voice of reason and compass for calmness throughout all of my studies.

To my parents Ivan and Adhika Naicker, thank you for the sacrifices you have made to support me up to and even beyond this point - this milestone is for you. Words cannot quantify the gratitude and appreciation I have for you both as you have been the foundation of my support structure throughout this journey that has been *ours*.

Thank you to my siblings Dr Nikita, Suhalia and Tharshey and my friends for guaranteeing I remain balanced throughtout this study.

Lastly, I acknowledge my supervisors Ilya and Francesco, you both have been integral in developing my skills for us to reach the end of this chapter. I appreciate your inputs and time spent with me over these years.

Romans 8:31: "What then shall we say to these things? If God is for us, who can be against us?"

Contents

List of Figures

List of Figures

List of Tables

List of Tables

Preface

The research discussed in this thesis is carried out in the College of Agriculture, Engineering and Science of the University of KwaZulu-Natal, Durban, South Africa from September 2019 until November 2022 by Kimara Naicker under the supervision of Prof. Francesco Petruccione and co-supervised by Prof. Ilya Sinayskiy.

I, Kimara Naicker, hereby declare that all the material incorporated in this dissertation are my own original work, except where acknowledgement is made by name or in the form of a reference. The work contained in herein has not been submitted in any form for any degree or diploma to any other institution.

Signed:............................ Date:.. 30 March 2023

University of KwaZulu-Natal, March 30, 2023

Abstract

This work focuses on using classical machine learning (ML) models to study the quantum dynamics of excitation energy transfer (EET) within strongly coupled open quantum systems relevant to light harvesting complexes (LHCs). Direct evidence for long-lived quantum coherence has been found to play an important role in EET processes during the first step of photosynthesis in certain LHCs where excitation energy is transmitted from the antenna pigments to the reaction center in which photochemical reactions are initiated [1–3]. The numerically exact method used to simulate the dynamics in this work is the hierarchical equations of motion (HEOM) adapted by Ishizaki and Fleming to suit the quantum biological regime [4–6]. In the case of an open quantum system, such as a photosynthetic pigment-protein complex, evolving over time we can generate a set of time dependent observables that depict the coherent movement of electronic excitations through the system by solving a suitable set of quantum dynamic equations such as the HEOM.

We have focused on solving two problems, the first being the inverse problem. That is, the objective is to determine whether a trained ML model can perform Hamiltonian tomography by using the time dependence of the observables as inputs. We demonstrate the capability of the convolutional neural network (CNN) to solve the inverse problem. That is, the trained CNN can accurately describe the system under study by predicting the parameters of the system Hamiltonian when given the aforementioned time dependent data. The models developed can predict Hamiltonian parameters such as excited state energies and inter-site couplings of a system up to 99.28% accuracy.

The second use of the same data set of observables involves time-series analysis. Although various analytical solutions for the dynamics of open quantum systems such as the HEOM have been developed, these often require immense computational resources. We demonstrate that models such as SARIMA, CatBoost, Prophet, convolutional and recurrent neural networks can predict the long-time dynamics provided that the initial short-time dynamics is given. Our results suggest that SARIMA can serve as a computationally inexpensive yet accurate way to predict long-time dynamics.

1

1   Introduction

Quantum mechanics is a fundamental theory in physics that describes the physical properties of nature at the scale of atoms and subatomic particles. Developments in observational techniques have allowed the study of the dynamics of biological systems on increasingly small scales. Some of these studies have revealed evidence of quantum mechanical effects which cannot be accounted for by classical physics in a range of biological processes. Quantum biology is the study of such processes. In most cases, physicists are interested in probing the evolution of such biological systems in the quantum regime. The simulation of the dynamics of quantum systems is one of the most challenging problems in physics and chemistry.

Machine learning can be seen as the art and science of computers learning how to solve problems from data instead of being explicitly programmed. During the past few years, machine learning models have revolutionised science and technology [7]. They have been used to classify images, translate between languages, control robots and self-driving cars and play complex games. In the context of physical sciences, machine learning models have been successfully used to predict the properties of materials, interpret astronomical pictures, classify phases of matter, represent quantum wave functions and control quantum devices [7–10]. Classical machine learning is further expected to play a substantial role in this rapidly growing inter-disciplinary field mainly owing to the availability of large amounts of data and of unprecedented computing power including the use of graphical processing units (GPUs); and we aim to contribute towards this through this work. There are many approaches to apply machine learning techniques to quantum biology and physics. Although still limited, much has already been done in the field to exploit the advantages that become available when merging such disciplines together. Based on existing quantum theory and data that can be constructed from such, we can design architectures of machine learning models which allow us to investigate the dynamics of biological systems - more specifically, photosynthetic complexes.

This thesis is a computational investigation into the potential of these models which gives some evidence that there is more to expect from artificial intelligence with physics. We have composed a conclusive literature review covering the research in the field of machine learning for physics in the following section. In recent years, there has been an exceptional rise of machine learning techniques impacting many fields, particularly in industrial applications including healthcare, finance and energy harvesting [7]. In parallel to this, scientists have become increasingly interested in the potential of machine learning for fundamental research and the field of physics is no exception.

The broad objective of machine learning is to recognize patterns in data which can then inform the

way unseen data samples are treated. To some extent there exist similarities between the disciplines of machine learning and physics with regards to their aims and methodology. Both involve the process of gathering and analysing data to inform models that are able to describe the behaviour of complex systems. However, the fields differ in the way their goals are realized. On the one hand, physicists aim to understand the mechanisms of nature. On the other hand, machine learning models are agnostic and the machine provides the 'intelligence' by interpolating the data. Therefore, machine learning tools in physics have been welcomed by some, while being eyed with suspicion by others.

Studying the exciton dynamics of large photosynthetic complexes has been a topic of growing theoretical and experimental interest in recent years [6, 11–19]. There are only a few fundamentally different natural light-harvesting complexes from which alone we can extract the relation between the structure of an excitonic system and its dynamics in full detail such as the Fenna-Matthews-Olson (FMO) and light-harvesting II complexes [20, 21]. There are diverse intersections of machine learning with physics in this regard and significant improvements on existing technologies have contributed to the success of machine learning in recent times. The theoretical society's understanding of the dynamics of these complexes is owed to theory papers that followed and were based on experimental findings which we will highlight in this section. Such tools can allow for detailed mechanistic studies of energy transfer processes in light-harvesting devices at the atomic level [7, 22].

Most of the developed machine learning models depend heavily on the knowledge of the Hamiltonian of a system as well as spectral densities in the case of describing the coupling of chromophores to their protein environments such as in a paper demonstrated by Valleau *et al.* [23].

The FMO complex was the first chlorophyll-containing protein that was crystallized [24] and it represents an important model in energy transfer and has been extensively studied experimentally and theoretically [1]. The site-basis Hamiltonian for the FMO system contains seven individual excitation energies along the diagonal and the dipole-dipole coupling terms in the off-diagonal positions. Estimating the Hamiltonian of a system is a difficult computational task due to the large number of degrees of freedom in these biological complexes. The first attempt using the extended dipole method with atomic coordinates obtained from the FMO crystallographic data, after the structure of the complex was solved by Fenna and Matthews in 1975 [20], was made by Pearlstein and Hemenger who calculated interchromophoric coupling energies [25]. Further attempts and improvements to the approximation of the FMO Hamiltonian since then involve traditional methods that exploit electronic spectroscopy [26]. On the other hand, machine learning techniques have been utilised for

---

[1]The Fenna-Matthews-Olson (FMO) complex is covered in greater detail in Section 3.1.

Hamiltonian tomography. The use of multi-layer perceptrons for predictions with regards to the FMO complex has been explored by Häse *et al.* [22]. Their methods were found to perform significantly faster than traditionally used time dependent density functional theory (TDDFT) approaches. These authors used multi-layer perceptrons as fully connected neural networks to predict the values of the first singlet excited state for the chromophores after training the neural networks on the excited state energies obtained from quantum mechanics and molecular mechanics calculations. Once the neural networks have made the excited state energy predictions, the Hamiltonian can be constructed from the predictions and the exciton dynamics can be computed. The 12 trained neural networks per bacteriochlorophyll predicted excited state energies with errors contained to 0.01 eV (0.5 %) from the neural network ensemble average. This study has further contributed to the demonstration of the power of machine learning in biochemistry.

Similar research has been carried out in recent work where machine learning was employed to extract other chemical properties [27–30] and potential energy surfaces [31]. These studies demonstrate that machine learning is proving to be a promising tool for enabling large scale excited state dynamics studies. Pigment-protein complexes exhibit remarkable transport properties which facilitate highly efficient excitation energy transfer across long distances [32–35]. As such, identifying working principles that ultimately transform into blueprints for novel nature-inspired excitonic devices is an active research frontier [36, 37].

Experimental studies reveal valuable insight into the microscopic details of excitation energy transfer. Prominent examples are given by studies probing the impact of electronic coherence or non-trivial interactions between excitons and specific vibrational modes on transfer characteristics [1, 38–49]. However, such investigations are tedious since they require sophisticated experimental setups, as well as computationally demanding simulations of open quantum system dynamics. In order to relate the dynamics to the underlying structure, it is desirable to investigate a large number of artificially designed excitonic systems [7]. This has been addressed in several other theoretical works [50–53].

In another study by Häse *et al.* [21], the authors have leveraged concepts from deep learning to overcome the great computational demand of established techniques for exploring energy transfer properties. They have trained multi-layer perceptrons here as well although in this instance, it is to predict overall transfer efficiencies and average exciton transfer times from an initially excited donor to a certain acceptor pigment. The input features to the multi-layer perceptrons correspond to the parameters of the corresponding Frenkel exciton Hamiltonians [54]. The trained neural networks evaluate transfer times on the order of milliseconds while also maintaining sufficiently high prediction accuracy. The presented results have been ratified by considering various artificial datasets

which were generated by uniform sampling of pigment excitation energies and inter-pigment couplings similar to the energies and couplings of a set of relevant biological complexes: the FMO complex as well as the light-harvesting complexes CP43, CP47 and the reaction centre of photosystem II. This work has revealed the potential of the multi-layer perceptrons to provide a significantly cheaper computational alternative to secular Redfield computations at comparable or even higher accuracy. Bhat *et al.* [55] developed a computational method in order to learn a molecular Hamiltonian matrix from matrix-valued time series of the electron density. Their sample set under study consisted of three small molecules. They proposed that the resulting Hamiltonians can be used for electron density evolution which produces highly accurate results up to 1 000 time steps beyond the training data. The learned Hamiltonians were further used to simulate electron dynamics in the presence of an applied electric field. This tests for generalization to a problem that is beyond the field-free training data. It was found that the resulting electron dynamics predicted by the learned Hamiltonian were in close quantitative agreement with the ground truth. The built method relies on combining a reduced-dimensional, linear, statistical model of the Hamiltonian with a time-discretization of the quantum Liouville equation within time-dependent Hartree Fock theory. The model was trained using a least-squares solver in a bid to avoid numerous CPU-intensive optimization steps. Much research has been conducted where machine learning has been applied to molecular simulation, these papers have been summarised in this paper by Bhat and co-authors [55].

The understanding of the fundamental processes of light-harvesting has proven to be valuable in the development of clean energy materials and devices. Biological organisms have evolved such that complex metabolic mechanisms can efficiently convert sunlight into chemical energy. This conversion has inspired the design of clean energy technologies including solar cells and photocatalytic water splitting [56].

Describing the emergence of macroscopic properties from microscopic processes poses the challenge of bridging length and time scales of several orders of magnitude. Machine learning has gained popularity as a tool to bridge the gap between multi-level theoretical models and trial-and-error approaches. Machine learning offers opportunities to gain detailed scientific insights into the underlying principles governing light-harvesting phenomena and can accelerate the fabrication of light-harvesting devices.

Molecular dynamics simulations with a purely machine learning based ground state density functional have recently been reported [57]. A study demonstrated how electron densities can be predicted from methodologies similar to kernel ridge regression and was used to evolve small molecular systems in their ground states over time [58].

The quantum mechanical simulations of charge and exciton transfer in molecular organic materials can further be seen as a key method to increase the understanding of organic semi-conductors. The goal of Krämer *et al.* [59] was to build an efficient multi-scale model to predict charge transfer movements and exciton diffusion constants from non-adiabatic molecular dynamics simulations and Marcus-based Monte Carlo approaches. In this paper, machine learning models were applied to simulate charge and exciton propagation in organic semi-conductors. In simulations, it was shown that kernel ridge regression models can be trained to predict electronic and excitonic couplings from semi-empirical density functional tight binding (DFTB) reference data with high accuracy. Using the developed models decreased the cost of exciton transfer simulations by one order of magnitude.

Electron transfer coupling is an important factor in determining electron transfer rates. This coupling strength can be sensitive to inter-molecular configurations, hence, studying the transportation behaviour of charge with a full first-principle approach comes with a large computational demand in quantum chemistry calculations. To tackle this issue, Wang *et al.* [60] developed a machine learning approach to evaluate electronic coupling. A machine learning model for an ethylene system was built by kernel ridge regression with Coulomb matrix representation. The performance of the machine learning model was found to be highly dependent on their building strategies which motivated the systematic investigation of the generalisation of the machine learning models. The optimal model configuration trained with 40 000 samples achieved a mean absolute error of 3.5 meV and greater than 98% accuracy in predicting phases and saved 10^4 times the computational cost of quantum chemistry calculations. The distance and orientation dependence of electronic coupling was successfully captured. This reinforces the notion that with the help of machine learning reliable charge transport models and mechanisms can be further developed.

Quantum chemistry calculations have proven to be useful in providing many key detailed properties and in enhancing the understanding of molecular systems. However, such calculations especially with ab initio models can be time consuming. For example, in the prediction of charge transfer properties, it is often necessary to work with an ensemble of different thermally populated structures. In another study by Wang *et al.* [61], an alternative to quantum chemistry calculations for the prediction of charge transfer properties by using a machine learning based approach is presented. In this work, the authors have shown that the prediction of electronic coupling, which is known to be sensitive to intermolecular degrees of freedom, can be obtained with artificial neural networks with improved performance as compared to the standard kernel ridge regression method.

Exact numerical simulations of dynamics of open quantum systems often require immense computational resources. Rodriguez *et al.* [62] developed a deep artificial neural network composed

of convolutional layers to simulate long time dynamics efficiently and accurately across different dynamical regimes from weakly damped coherent motion to incoherent relaxation. They were able to demonstrate that the model is a powerful tool for predicting long time dynamics of open quantum systems provided the preceding short time evolution of a system is known. Their model was trained on a data set relevant to photosynthetic excitation energy transfer to be applied to study quantum coherence occurrences observed in light-harvesting complexes. Moreover, the model is able to generalize well on initial conditions that are different to those used in training. Their approach reinforces the promise neural networks hold for becoming a valuable tool in the study of open quantum systems.

The modelling of non-adiabatic, long time dynamics in complex systems is known to be challenging. In this research paper by Wu *et al.* [63], a time series machine learning architecture including the hybrid convolutional neural network and long short term memory (CNN-LSTM) framework is proposed for predicting the long time quantum behaviour when given only the short time dynamics. This model leverages the local feature extraction capability of CNNs and the long term global sequential pattern recognition ability of LSTM. The developed hybrid model is shown to perform well in predicting the linearized semi-classical and symmetrical quasi-classical mapping dynamics as well as the mixed quantum-classical Liouville dynamics of various spin-boson models with learning time up to 0.3 ps. The hybrid CNN-LSTM network is believed to have high predictive power in forecasting the non-adiabatic dynamics in realistic charge and energy transfer processes in photo-induced energy conversion. Furthermore, if the hybrid network has learned the dynamics of a system, this knowledge is transferable that could significantly enhance the accuracy in predicting the quantum population and coherence dynamics of a similar system.

In a study by Lin *et al.* [64], recurrent neural networks with the long short-term memory cell (LSTM-NN) are employed to simulate the long time dynamics of open quantum systems. For the construction and prediction of the LSTM network, the bootstrap method is applied which offers a Monte Carlo estimation of forecasting confidence interval. A large number of LSTM networks are constructed by re-sampling time series sequences that were obtained from the early stage quantum evolution given by the numerically exact multi-layer, multi-configurational and time-dependent Hartree method. The constructed LSTM network ensemble is used for the propagation of the long time quantum dynamics and the simulated result is found to be in agreement with the ground truth evolution. The forecasting uncertainty that partially reflects the reliability of the network prediction is considerably low. This work exhibits the practicality and capability of the bootstrap based LSTM approach to propagate the long time quantum dynamics of open systems with high accuracy and low computational cost.

7

Two-dimensional electronic spectroscopy is one of the experimental tools for exploring the dynamics of excitonic energy transfer in biological complexes. With the objective of extracting model parameters from experimental spectral data, Rodriguez *et al.* [65] show how computationally expensive but exact theoretical methods encoded into an artificial neural network are used to extract model parameters and extrapolate dipole orientation from two-dimensional electronic spectra (2DES) or reversely i.e. to produce 2DES from model parameters.

The central theme focused on in this thesis is that of applying machine learning techniques to the description of open quantum system dynamics, particularly the excitation energy transfer dynamics within the FMO light-harvesting complex. Within the supervised learning regime of machine learning, statistical models and artificial neural network models prove to be purposeful in reaching our aims of better understanding the role of quantum mechanics in the movement of converted energy within the FMO and other biological, photosynthetic complexes and, of attempting to bypass or reduce the demand of computational resources that are required to perform simulations with conventional methodology.

Following this introductory section, Chapters 2, 3, 4 and 5 provide the necessary theoretical foundations for understanding the subsequent chapters which highlight the main contributions of the thesis by the authors. Chapter 2 presents concepts and methods from the theory of open quantum systems, Chapter 3 presents concepts and methods from quantum biology, Chapter 4 presents the theoretical framework of the hierarchical equations of motion and Chapter 5 presents concepts and methods from machine learning. Chapter 6 and Chapter 7 present the main contributions of this study. All of this is followed by Chapter 8 containing concluding remarks.

2 Open Quantum Systems

A reduced open quantum system is by definition "what we are interested in", however, to investigate its properties we need to know the effects of a thermal bath on the relevant system, but the bath has to be treated statistically due to its size or complexity. Within the Theory of Open Quantum Systems, there are many sub-categories, beyond simply quantum or classical, for the types of environments that can be considered. Within the quantum environment category there are bosonic or fermionic environments, stationary or non-stationary environments and within the stationary type there are thermal (often called reservoirs) and non-thermal environments.

As a result of the system-environment coupling, the system also influences the environment and possibly alters its evolution. However, in most circumstances these changes induced by the system into the environment are not significant, due to the great size and energy of the environment in comparison to the system, and can usually be described as stationary. For the case of a quantum system in contact with a reservoir, as often encountered in thermodynamics and statistical physics, if it is prepared in a given quantum state then it will evolve due to the presence of the reservoir until it reaches thermal equilibrium with the reservoir.

The presence of the environment is commonly seen as the largest contributing factor to the destruction of quantum features within a system. Basically, the environment introduces noise into quantum systems that destroy delicate coherent superpositions, thereby, transforming them into classical statistical mixtures.

2.1 Quantum mechanics

Quantum mechanics is a theoretical framework that provides a description of an isolated physical system by a quantum state that completely characterizes the system that is denoted by $|\psi\rangle$ and a Hilbert space $\mathcal{H}$ that contains all the possible states available of the system [66]. There are four main postulates on which Quantum Mechanics can be built.

2.1.1 Hilbert space

Postulate 1: Associated with any isolated physical system is a complete complex vector space with inner product (a Hilbert space) known as the state space of the system. The system is completely described by its state vector, which is a unit vector in the system's state space.

A Hilbert space is a vector space equipped with an inner product. The postulate means that physical states of a quantum system can be associated with a vector $|v\rangle \in \mathcal{H}$.

The linear combination or superposition of multiple states constituting the Hilbert space can also be used to describe the system. The existence of superposition states in the quantum regime results in unique quantum-mechanical properties which are often counter-intuitive when compared to classical mechanics.

Quantum coherence describes the relationship between the individual states forming a superposition and quantum entanglement is a form of correlation between quantum states.

2.1.2 Combining Quantum systems

Postulate 2: Given two quantum systems with respective Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$ the combined quantum system has associated with it a Hilbert space given by $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$.

2.1.3 Time evolution

The evolution of a closed quantum system is described by a unitary transformation. The time evolution of a system is the main factor used to understand its nature and properties. When dealing with classical systems, time evolution is usually formulated in terms of differential equations. Erwin Schrödinger obtained the first quantum evolution equation in 1926 [67] when physicists were often translating methods which were useful in the classical regime to the quantum case. Formally known as Schrödinger's equation, it describes the behaviour of an isolated or closed quantum system which is defined as a system that does not interchange information such as energy or matter with another system.

Postulate 3: The state vector of the system obeys the Schrödinger equation,

$$i\hbar \frac{d}{dt}|\psi\rangle = H|\psi\rangle, \tag{1}$$

where $\hbar$ is the reduced Planck constant, here we consider units where $c = \hbar = 1$, and H is the Hamiltonian, which represents the energy levels of the system and the interaction between parts of the system.

The solution of the Schrödinger equation can be presented in the form of a unitary operator $U(t)$ such that the time evolution of a state $|\psi\rangle$ is given by

$$|\psi(t)\rangle = U(t)|\psi(0)\rangle. \tag{2}$$

For a closed quantum system where H is independent of time, it follows that $U(t) = exp(-iHt)$. If we know the Hamiltonian of a system then at least in principle, we understand its dynamics completely.

2.1.4   Measurements

Postulate 4: Quantum measurements are described by a set M_k of measurement operators satisfying the constraint $\sum_k M_k^\dagger M_k = I$. These are operators acting on the state space of the system being measured. A state $|\psi\rangle \in \mathcal{H}$ becomes $\frac{M_k|\psi\rangle}{\sqrt{p_k}}$ instantaneously after the measurement with probability $p_k = ||M_k|\psi\rangle||^2 \geq 0$.

The mathematical apparatus to calculate the probabilities and expectation values of measurement outcomes for a quantum system is largely based on linear algebra calculus.

2.2   The density operator

The study of density operators is motivated by considering ensembles of pure quantum states that measurements are made on. Suppose we have a system in state $|\psi_1\rangle$ with probability q_1 or in state $|\psi_2\rangle$ with probability q_2 and so on. Then we have a pure state ensemble $\{q_i, |\psi_i\rangle\}_{i=1}^N$ to describe the system. We consider measurements made on this quantum system. Suppose the system is in state $|\psi_i\rangle$ and it is measured with a set of measurement operators M_k, the measurement transformation is

$$|\psi_i\rangle \rightarrow \frac{M_k|\psi_i\rangle}{\sqrt{p_{k|i}}} = |\psi_i^k\rangle \tag{3}$$

with probability $p_{k|i} = \langle \psi_i|M_k^\dagger M_k|\psi_i\rangle$ which gives the probability of outcome k given state $|\psi_i\rangle$.

Consider the case where the state is unknown but it is known that the state belongs to the ensemble $\{q_i, |\psi_i\rangle\}_{i=1}^N$ then the probability of outcome k after measurement on the ensemble is

$$p_k = \sum_k p_{k|i} q_i = \sum_i q_i \langle \psi_i|M_k^\dagger M_k|\psi_i\rangle = \text{Tr}\left[M_k^\dagger M_k \left(\sum_i q_i|\psi_i\rangle\langle\psi_i|\right)\right]. \tag{4}$$

Quantum theory can also be formulated in terms of the outer product of a vector in $\mathcal{H}$, a density operator $\rho = |\psi\rangle\langle\psi|$, which is a unit-trace Hermitian positive operator. A density operator allows us to consider an ensemble or mixture of quantum states, $\rho = \sum_k p_k|\psi_k\rangle\langle\psi_k|$ with $\sum_k p_k = 1$ and p_k being classical probabilities of the system being in state $|\psi_k\rangle$. The density matrix is completely equivalent to the pure state ensemble. Using the density matrix, Eq. (4) becomes $p_k = \text{Tr}(M_k^\dagger M_k \rho)$.

Now we may look at the states after measurement. If outcome k is observed for initial state $\rho = \sum_i q_i|\psi_i\rangle\langle\psi_i|$ then using Eq. (3), $\{p_{k|i}, |\psi_i^k\rangle\}_{i=1}$ is the resulting ensemble where $p_{k|i}$ is the probability of outcome k given state $|\psi_i\rangle$. In the alternate case where outcome k is observed but the initial state is unknown then we should sum over all possible states with outcome k with their respective probabilities $p_{i|k}$ where the the density operator for outcome k is

$$p_k = \sum_i p_{i|k} |\psi_i^k\rangle\langle\psi_i^k| = \sum_i p_{i|k} \frac{M_k|\psi_i^k\rangle\langle\psi_i^k|M_k^\dagger}{p_{k|i}} = \sum_i \frac{q_i}{p_k} M_k|\psi_i^k\rangle\langle\psi_i^k|M_k^\dagger = M_k\rho M_k^\dagger = \frac{M_k\rho M_k^\dagger}{\mathrm{Tr}[\rho M_k M_k^\dagger]}.$$

(5)

Therefore, we can compare the pure state case with the generalised density operator case as

$$|\psi\rangle \to \frac{M_k|\psi\rangle}{\sqrt{p_k}} \quad \text{and} \quad \rho \to \frac{M_k\rho M_k^\dagger}{p_k}.$$

(6)

Let us now consider an example. A qubit is a quantum state $|\psi_i\rangle$ in a two-dimensional Hilbert space $\mathcal{H} = \mathbb{C}^2 = \mathrm{span}\,\{|0\rangle, |1\rangle\}$ where $|0\rangle$ and $|1\rangle$ form an orthonormal basis for $\mathcal{H}$. The density operator for any state in this space is thus of the form $\sum_i q_i |\psi_i\rangle\langle\psi_i|$ and can hence be represented by a 2×2 complex matrix of the form

$$\rho = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

(7)

However, the unit trace property reduces to the condition $d = 1 - a$ and Hermiticity property reduces to the condition that $c = b^*$ and that a be real. Thus, the density matrix is completely parametrized by the complex number b and the real number a and takes the form

$$\rho = \begin{pmatrix} a & b \\ b^* & 1 - a \end{pmatrix}.$$

(8)

A pair of basis vectors $\vec{u}$ and $\vec{v}$ are orthogonal when their inner product $\vec{u}^\mathsf{T} \cdot \vec{v}$ is zero. The following Pauli matrices form an orthogonal basis for the 2D space

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

(9)

Any qubit density matrix can represented by

$$\rho = \frac{1}{2}(I + \vec{v} \cdot \vec{\sigma}) = \frac{1}{2}\begin{pmatrix} 1 + v_z & v_x - iv_y \\ v_x + iv_y & 1 - v_z \end{pmatrix},$$

(10)

where $\vec{v} = (v_x, v_y, v_z)$ and $\vec{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$. $\vec{v}$ is the *Bloch vector*. In the case where positivity of the density matrix is required, the constraint $|\vec{v}| \leq 1$ is applied to the Bloch vector. For any unit Bloch vector $\text{Tr}\rho^2 = 1$ which represents a pure state and Bloch vectors of length less than 1 yield mixed states. The set of all valid Bloch vectors $\vec{v} = \text{Tr}(\rho\vec{\sigma})$ can be represented by the *Bloch sphere*. The Bloch sphere can describe all qubit states and can be embedded in three dimensions, hence, it is a useful tool for illustrating various qubit states, although it is typically only used to represent two-level systems as shown in Figure 1.
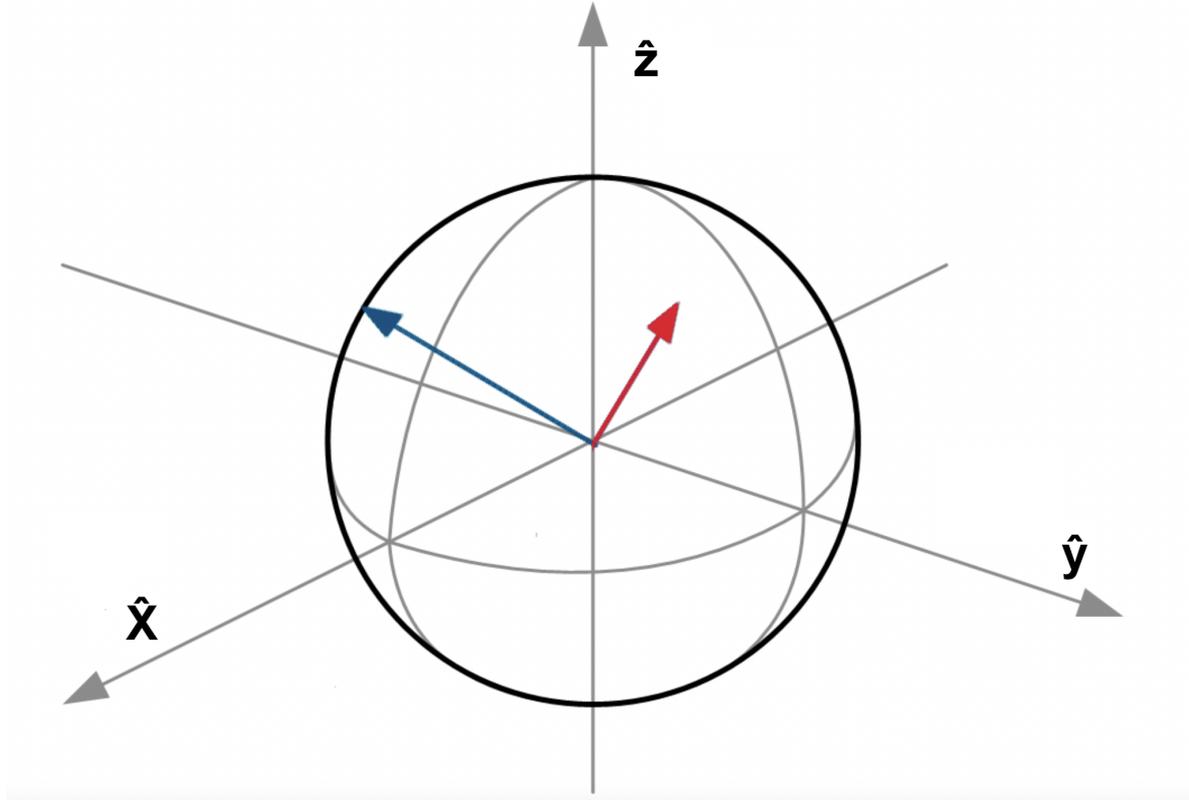
Fig. 1: The Bloch sphere is a geometric representation of the collection of all Bloch vectors which describe valid qubit density operators. Therefore, the sphere is of radius 1, its surface represents all pure states and its interior represents all mixed states. In this diagram, the blue vector lies on the surface of the sphere indicating a pure state whereas, the red vector lies in its interior indicating a mixed state.

2.3 Composite systems

Now that we have discussed in detail the dynamics of a single system, consider a two component system where we have a sub-system of interest A (often simply referred to as the system) and the other sub-system is the bath B. By Postulate 3, we assume that the total system is described by a density matrix ρ_t, and it evolves according to the Schrödinger equation. By Postulate 2, the Hilbert space of the combined system is the tensor product of the individual spaces $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$.

The partial trace is used to obtain the description of a sub-system from a composed system. The partial trace is a linear operator that maps from the total Hilbert space to the Hilbert space of A. It establishes the relation between the total density matrix $\rho_{AB} \in \mathcal{H}$ and the density matrix of the sub-system A ρ_A.

In Dirac notation the trace over sub-system A of a joint quantum system $\mathcal{H}_B$ with basis $|\psi_{bi}\rangle$ is

$$\rho_A = \mathrm{Tr}_B \rho_{AB} = \sum_i \langle \psi_{bi} | \rho_{AB} | \psi_{bi} \rangle. \tag{11}$$

2.4 Time evolution in open quantum systems

In quantum mechanics the pure state of a system can be described in terms of vectors in Hilbert space. However, if there is statistical mixture of pure states, that is, if the final state had been prepared in one of many possible pure states randomly chosen according to some probability, then the density operator is used for the representation of the statistical mixtures or for pure states as well. This mathematical object is useful for understanding the correlations between an open quantum system (OQS) and its environment. To describe the evolution of a system, the equivalent of Schrödinger's equation in the density-operator formalism which is the von Neumann equation of motion can be solved for the temporal evolution of the total state of the system and the environment. The state of the system and the environment together is considered closed, hence, remains in a pure state at all times.

The state of the system at any point in time can then be determined by tracing out the environment. Mathematically, this refers to the integration over the degrees of freedom of the environment to compute the reduced state of the system. This produces the partial trace or the reduced density matrix of the OQS.

The caveat, however, is that solving for the total state of system plus environment is generally unfeasible and often deems this approach impractical. A goal of the theory of OQS is to avoid having to integrate the full system. Alternatively, by analysing the Hamiltonian, an approximate equation of motion for the system's density operator can be constructed. This equation of motion is called a master equation. There are several approaches to the dynamics of open quantum systems. Each description presents a departure from the standard Liouville-von Neumann equation

$$i \frac{d}{dt} \rho(t) = [H, \rho(t)], \tag{12}$$

which governs the evolution of the closed system fully characterized by the system Hamiltonian H.

The structure of the system-environment initial state is essential for determining the nature of the evolution of the reduced density matrix of the OQS, defined as $\rho_S(t) = \mathrm{Tr}_B\{\rho_{Tot}(t)\}$, where $\rho_{Tot}(t)$

is the density operator of the full system.

The approach taken to study the dynamics of an open quantum system differs depending on the specific system being analysed [54, 68–71]. Fundamental properties such as the concept of Markovianity have often proven not to be straightforward, hence, background for this work is therefore reduced to familiarity with the basics of quantum mechanics.

2.5   Non-Markovianity

Typically, the dynamics of an OQS can be described with a Markov approximation which assumes that the environment recovers instantly from the system-environment interaction, therefore, allowing a continuous flow of information from the system to the environment. To be Markovian, it is assumed that when energy is transferred from the system to the environment during their interaction, the energy is "instantly" absorbed into the environment and, if energy is transferred from the environment to the system at a later stage then this later quantum of energy is unrelated to the initial transfer (so called white noise). Within the Markov approximation, master equations can often be arranged in the well-known Gorini–Kossakowski–Sudarshan–Lindblad (GKSL) form [72, 73]. The most used approaches for the kind of problems, where assumptions are made that the system-bath interaction is weak (perturbative treatment) and that the correlation time of the bath fluctuations is very short (Markovian assumption), are the Redfield equation [74] and the quantum master equation [75, 76].

The key idea is that there must be a clean separation between the typical correlation time of the fluctuations and the time scale of the evolution that we want to follow. Crudely speaking, we may denote by δt_E the time it takes for the bath to 'forget' information that it acquired from the system. After time δt_E we can regard that information as forever lost and neglect the possibility that the information may feed back again to influence the subsequent evolution of the system.

The GKSL equation is the most general form of a Markovian master equation. We will follow a simple deduction presented by H. P. Breuer and F. Petruccione [77] where the Born-Markov master equation is derived from the Liouville-von Neumann equation and we see that the Lindbladian is produced as a result of the Markov approximation. It is shown that each Lindblad represents the effect on the system caused by the system-bath interaction.

We initially consider a system S whose dynamics is the subject of interest. The system interacts with a bath B and this interaction is represented by SB. The total Hamiltonian is

$$H_{Tot} = H_S + H_B + H_{SB}, \tag{13}$$

where H_S, H_B and H_{SB} are Hamiltonians for the system, bath and system-bath interaction, respectively. The Liouville-von Neumann equation for the total density operator ρ_{SB} is

$$\frac{d}{dt}\rho_{Tot} = -i[H_{Tot}, \rho_{Tot}]. \tag{14}$$

In the interaction picture, the Hamiltonian and density operator can be written respectively as,

$$H_{SB}(t) = e^{\frac{i}{\hbar}(H_S+H_B)t} H_{SB} e^{-\frac{i}{\hbar}(H_S+H_B)t}, \tag{15}$$

and

$$\rho_{Tot}^I(t) = e^{\frac{i}{\hbar}(H_S+H_B)t} \rho_{Tot} e^{-\frac{i}{\hbar}(H_S+H_B)t}. \tag{16}$$

The equation for the time evolution for the density operator $\rho(t)$ is

$$\frac{d}{dt}\rho_{Tot}^I(t) = -i[H_{SB}(t), \rho_{Tot}^I(t)]. \tag{17}$$

The time evolution of the reduced density matrix $\rho_S(t)$ is given by

$$\rho_S(t) = \text{Tr}_B\{\rho_{Tot}(t)\} = \text{Tr}_B\{e^{\frac{i}{\hbar}(H_S+H_B)t} \rho_{Tot}^I e^{-\frac{i}{\hbar}(H_S+H_B)t}\}. \tag{18}$$

Using the iterative approach, together with the Born and Markov approximations, one can show that the Liouville equation becomes a quantum master equation in the Born-Markov form

$$\frac{d}{dt}\rho_S(t) = -\frac{1}{\hbar^2} \int_0^\infty dt' \text{Tr}_B[H_{SB}(t), [H_{SB}(t-t'), \rho_S(t)\rho_B]]. \tag{19}$$

It is well-known that the above equation together with microscopic model of the system-bath interaction and rotating wave approximation reduces to the quantum master equation in the Lindblad form

$$\frac{d}{dt}\rho_S(t) = -i[H_S, \rho_S] + \mathcal{L}(\rho_S), \tag{20}$$

where $\mathcal{L}(\rho_S)$ is the Lindblad dissipater $\mathcal{L}(\rho_S) = \sum_k \gamma_k(A_k\rho_S A_k^\dagger - \frac{1}{2}A_k^\dagger A_k, \rho_S)$, where γ_k are non-negative constants and A_k are operators acting on the reduced density matrix [77].

It is important to note that these equations do not satisfy the necessary positivity condition without the imposition of a rotating wave approximation or a secular approximation [78–80]. The caveat is that such approximations alter the form of the system-bath interaction [81] and the dynamics of the original total Hamiltonian [82, 83].

For the type of environment that is described as a set of independent harmonic oscillators, the system-environment interaction is characterized by a spectral density and correlation function which fully captures the action of the environment on the OQS dynamics. Such an environment has an infinite

number of degrees of freedom and is, at least initially, in a thermal equilibrium state. The behaviour of the open system depends on the correlation function $\alpha(t)$ which is determined by the shape of the spectral density function $J(\omega)$. The time scale of the decay of the correlation function defines the environmental relaxation time which corresponds to the time that the environment takes to return to its initial state. If this time scale is much smaller than the evolution time of the system then, a Markovian approximation shall be considered to derive the OQS evolution equations.

It is commonly found that when a large separation between the system and environment time scales can no longer be assumed, non-Markovian behaviour is observed and a back-flow of information from the environment into the system occurs. In many chemical physics problems, the environment is complex and strongly coupled to the system at finite temperature. Consequently, the Markovian approximation, rotating wave approximation and perturbative expansions become invalid [82]. Research areas such as energy transport in photosynthetic systems [2, 6, 37, 84–86], quantum thermodynamics [87, 88] and the strong coupling regime in artificial light-matter systems [89–92] have demanded the development of numerically exact methods to explore non-perturbative and non-Markovian regimes. Following this, a great amount of research has been dedicated to studying the problems of OQS dynamics for the development of an accurate but efficient descriptions beyond the Markov approximation [93–95].

The modified Redfield [96] and time-convolution-less Redfield equations [97] are reduced equations of motion. Although these approximate approaches are useful for studying problems of OQS dynamics, their range of validity is restricted.

The pseudomode [98] and the reaction coordinate mapping approaches [99] consider equations of motion that utilize an effective mode whose dynamics are described by the Markovian master equation. While these approaches have a wider range of applicability than the conventional reduced equation of motion approaches, the description of long-time behaviour that they provide might be unreliable as the Markovian master equation cannot predict the correct thermal equilibrium state (particularly in low temperature cases). The introduction of a pseudo-Matsubara mode can be used to overcome this limitation as in the case of the hierarchical equations of motion formalism [100].

The variational approaches mentioned above can be used to treat non-linear system-bath coupling but since the bath is described as a finite number of oscillators, the number of bath modes must be increased until convergence is achieved to obtain more accurate results. This means that the study of long-time behaviour using these approaches requires a demanding computational effort. On the other hand, a reduced equation of motion approach requires a numerical effort that scales only linearly with the simulation time.

Over time, numerically 'exact' approaches have been developed that are not subject to the restraints of the above variational approaches. Here, numerically exact suggests the capacity to calculate the dynamical and thermal properties of a reduced system with any preferred accuracy that can be clearly confirmed through non-Markovian tests based on exact analytical solutions. Early research focused on the problem of a quantum system interacting with an environment in the non-Markovian regime is highlighted in Bretón *et al.* [101] and Mukamel *et al.* [102].

3   Quantum Biology

The role of quantum effects and its contribution to the understanding of biological systems is a controversial topic. "Quantum biology" is the application of quantum theory to aspects of biology where classical physics typically fails to give an accurate description. The origin of Quantum biology is often traced back to 1944 and the publication of Erwin Schrödinger's famous book, 'What is Life?' [103]. However, even before this, the German physicist Pascual Jordan published a book a year before Schrödinger's, entitled Physics and the Secret of Organic Life [104] where he posed the question 'Sind die Gesetze der Atomphysik und Quantenphysik für die Lebensvorgänge von wesentlicher Bedeutung?' ('Are the laws of atomic and quantum physics of essential importance for life?'). Jordan also had been using the term 'Quantenbiologie' ('Quantum biology') since the late 1930s. The main source of controversy around the topic is the existence of decoherence. This is because quantum effects are delicate and on the other hand, biological systems naturally and unavoidably interact with their environments thereby creating favourable conditions for decoherence to occur. Although there exists experimental evidence to support the existence of quantum effects in olfaction [105] and magnetoreception [106], it is the coherent energy transfer in photosynthesis that is the most experimentally successful case.

The transport of energy between pigments in biological photosynthetic complexes is affected by the environment produced by surrounding vibrating proteins [107]. Experiments have proven the existence of long-lasting coherences between excitons (pigments) in several types of photosynthetic complexes, even at physiological temperatures [1, 3, 44]. As a result, pigments involved in this energy transport could in principle be considered as an open quantum system coupled to the surrounding environment [15, 108–110].

The relaxation time of this environment in a typical situation can be comparable to or slower than the rate of electronic energy transfer dynamics within the pigment complex, therefore, a Markov approximation is no longer accurate [111, 112]. This led to the dynamics of these systems being studied beyond the Markov approximation which allows for the consideration of the full system dynamics (see Section IX of [93]). Other common approaches, such as the hierarchy approach (see Section VI.C of [93]), are based on calculating the reduced dynamics of the system under the assumption that the environment evolves much faster than the open quantum system itself. Consequently, the environmental degrees of freedom can be traced out. Lambert *et al.* [84] provides an excellent review on quantum biology.

19

3.1 Fenna-Matthews-Olson complex

There is one pigment-protein complex, the so-called Fenna-Matthews-Olson (FMO) complex of green sulfur bacteria, which has been extensively used as a model system for larger antenna complexes, starting more than 40 years ago with the pioneering work of Pearlstein and co-authors [25]. The FMO complex of *Prosthecochloris aestuarii* was the first pigment-protein complex for which the structure was determined using X-ray crystallography [20, 113]. Since the resolution of the electron density map was refined [114] and the structure of the FMO complex of a strongly related bacterium *Chlorobium tepidum* was determined [115], it has been used for many years to explore structure-function relationships in antenna proteins [116]. The two structures are very similar, but interestingly, the spectra look different [11].

The FMO pigment-protein complex provides a paradigmatic model system in terms of the observed quantum coherence associated with photosynthetic energy transfer processes [1, 6, 45, 116]. Quantum coherence in this sense is electronic coherence i.e. the presence of a quantum mechanical superposition of the energetic eigenstates of the photosynthetic transport system; these terms are used interchangeably in the text. In green sulfur bacteria, the FMO complex conducts energy from the light-harvesting antenna to the reaction centre where it is used to trigger further processes to bind its energy in a chemical form [117, 118]. Each of the three non-interacting FMO monomers contains seven coupled bacteriochlorophyll-*a* chromophores arranged asymmetrically, yielding seven non-degenerate, delocalized molecular excited states called excitons [20, 45]. It is this small number of states that makes this complex attractive for theoretical studies of energy transport dynamics. A graphical representation of the FMO complex by Huo *et al.* [119] is given in Figure 2.

As shown by Ishizaki et al. [6], the arrangement of the chromophores in FMO complex results in a downhill and uneven energetic landscape with two different routes through which an excitation can be transferred to reach the lowest energy state. It is important to note that the lowest energy state is on the pigment closest to the reaction centre, so that the excitation energy can be rapidly transported to the reaction centre.

Investigations of photosynthetic systems carried out at 77 K have found evidence of coherent energy transfer in many antenna complexes [1] and even in the reaction centre of purple bacteria [3, 120, 121]. In the context of two-dimensional electronic spectra (2DES) used in these investigations mentioned, electronic coherence is visualized by oscillations of certain off-diagonal peaks of the 2D spectrum with the population time. It is important to note that this wave-like energy transfer mechanism is only able to contribute to the near perfect quantum efficiency of
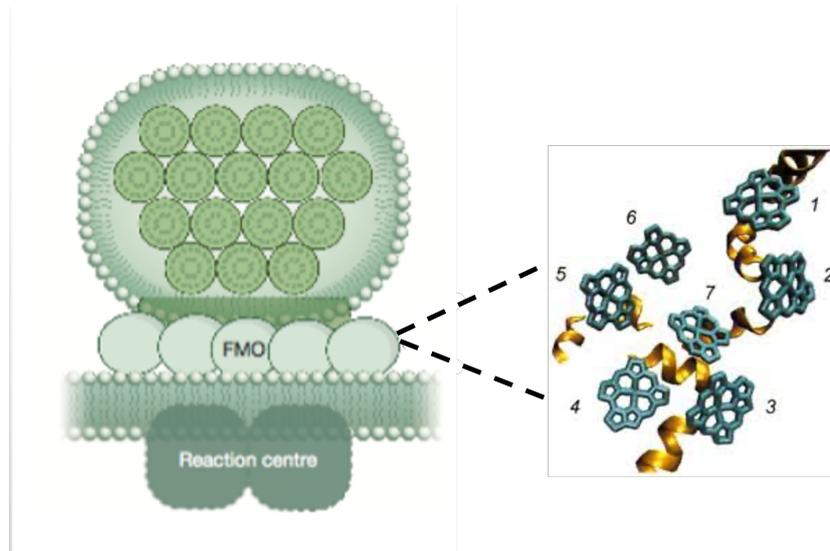
Fig. 2: The Fenna-Matthews-Olson pigment-protein complex serves as a bridging energy wire as it is tasked with transporting energy hv in the form of light harvested in the chlorosome antenna to the reaction centre pigments [13, 20, 37, 119].

photosynthesis if the coherence survives in these systems during energy transfer at physiological temperatures. It has been found that as temperature increases, thermally excited vibrational modes of the protein environment cause larger energetic fluctuations, thus accelerating the dephasing of coherence [3, 117, 122].

Although the decoherence that occurs seems unfavourable, research has been conducted showing that the delicate interplay between quantum coherence and dephasing can create transfer pathways within these complexes that result in highly efficient electronic energy transfer [15, 123]. Though, this set-up exploits quantum coherence to overcome an energy barrier, subsequent dephasing processes trap the excitation at the target site.

This suggests that it is the constructive interplay between dephasing noise and coherent dynamics, as opposed to the mere presence or absence of entanglement, that is responsible for the improved transfer of excitations [15]. It is intriguing that Nature appears to exploit noise-assisted processes to maximize a system's performance.

Although other ultrafast methods in the 1990s and early 2000s were the first to reveal the existence of quantum fluctuations [54, 117], ultrafast transient absorption spectroscopy methods have been the ideal standards for several decades to reveal the details of excitonic energy transfer in photosynthetic complexes [124]. Experimentally, quantum coherences in the FMO complex can alternatively be probed using two-dimensional electronic spectroscopy to directly observe electronic couplings and quantum coherences as a function of time [1]. The experimental method and theory have been

described in great detail by Brixner *et al.* [125].

The event where the initial excitation or transfer takes place initiates quantum coherence because both the dipole and site operators do not commute with the Hamiltonian of the system. The diagonal terms represent populations in the excitonic basis and the off-diagonal terms describe coherences. The phase factors in the coherence terms are responsible for quantum beating. This beating performs as a periodic modulation of peak amplitude and population in the site basis. The frequency of this beating corresponds to the energy difference between the two excitons giving rise to that specific quantum coherence.

It is commonly found that this phenomenon is ignored in transport dynamics because fast electronic dephasing generally destroys quantum coherence before it can impact the transport process.

As an example, at cryogenic temperature, the electronic coherences between the ground and the excited states in FMO complex are found to dephase in approximately 70 fs [3]. In contrast, electronic coherences among excited states have been shown to persist past 660 fs which is considered long enough to improve transport efficiency [1].

Such an electronic coherence persists if the electronic spectral motion among chromophores is strongly correlated, a demonstration of such is shown by Collini and Scholes [126]. These correlations tend to arise because the protein environment forces transition energies of chromophores to fluctuate together due to spatial uniformity of the dielectric bath. The result being long-lived quantum coherences that evolve in time, which in turn create periodic oscillations in both spectral signals and wavepacket position. This quantum beat provides the signature of quantum coherence [3].

Based on these experimental observations [127], it has been suggested that the enhanced energy transfer rates may be attributed to the exploitation of principles of quantum search algorithms by the quantum dynamics of the FMO complex [1, 128]. There is already an important body of work analysing energy transport in these systems that has been collated by Caruso *et al.* [108].

Further theoretical studies of dephasing the process of exciton transport in light-harvesting complexes have been carried out following these experimental results and their interpretations [14, 15]. Contrary to the conventional intuition that noise always deteriorates the performance of a system, in several classical as well as quantum instances, noise is known to be of advantage in enhancing a system's response. In the classical occurrence of stochastic resonance, a non-linear system can benefit from the presence of noise to achieve an enhanced sensitivity to a weak signal. This observation motivated the experiment which successfully demonstrated that noise may have the capacity to generate genuinely quantum mechanical features such as quantum coherence and entanglement [122, 129, 130]. It has

been further shown that local dephasing can enhance the rate and efficiency of energy transfer when compared to a perfectly quantum coherent system [14, 15]. Undeniably, the exciton transfer in light-harvesting complexes requires the concurrent presence of quantum coherent evolution and dephasing noise.

Caruso *et al.* [108] describe fundamental mechanisms underlying dephasing-assisted transport and elucidate them through numerical simulations. They show that after approximately 1 ps, the system evolves into a pure state where the main population is roughly evenly distributed amongst sites 1 and 2. Here, the amplitudes have a relative phase close to π, thereby leading to destructive interference for transitions to site 3. The destructive interference must be disrupted by a moderate level of dephasing noise on the sites 1 and 2. Moreover, it is critical to note that the energy of site 3 differs from that of sites 1 and 2 to an extent that the Hamiltonian matrix elements between sites 1 and 2 are insufficient to cause strong transfer to site 3. For this reason, a modest amount of dephasing on sites 1, 2 or 3 may assist the transfer process by means of line broadening to strengthen this transfer channel. It is undesirable for the system to transfer population into the sites 4, 5, 6 and 7 as this will prolong the time until transfer into the reaction centre through site 3. Hence, strong dephasing on those sites would be ideal to overpower unwanted exploration of them. As both sites 1 and 2 are most strongly coupled to site 6, the strongest dephasing would be expected for site 6 out of sites 4, 5, 6 and 7.

The basic mechanisms described above suggest a moderate dependency on the noise level as only vanishing or exceptionally strong dephasing noise led to a decline in performance. The optimal dephasing values were found to be resilient under variations in the system parameters such as site energies, coupling strengths and dephasing rates which may imply that the experimental results recorded for samples at 77 K could also be observed at higher temperatures.

Olaya-Castro *et al.* [127] inspect the function of quantum coherence in the efficiency of exciton transfer in a ring-hub arrangement of interacting two-level systems. This architecture imitates a light-harvesting antenna connected to a reaction centre. They found that the efficiency of excitation transfer from the antenna to the reaction centre is sensitive to the quantum superposition properties of the initial state, such as symmetric and asymmetric superposition of states in the basis of localized excitations, indicating that initial state properties can be used as an efficiency control parameter at low temperatures.

Rebentrost *et al.* [123] also demonstrated that exciton transport efficiency can be improved with pure dephasing induced by a fluctuating environment. This contrasts with fully coherent hopping that leads to localization in disordered systems and with highly incoherent transfer that is eventually suppressed by the quantum Zeno effect. They support that quantum transport can also be affected by quantum

localization.

In chromophoric complexes, an environment-assisted quantum walk approach within the Lindblad model involving relaxation and dephasing was an alternative suggestion used to explain the high-energy transfer efficiency [14]. This approach was also used to quantify the percentage contributions of quantum coherence and environment-induced relaxation to the overall efficiency [109].

3.2 Energy transfer

The pigments in the FMO complex are held by a protein scaffold at the right distances and orientations for efficient excitation energy transfer. A valid question might be - how does the protein environment influence the optical transition energies of the pigments? Besides static and dynamic disorder caused by slow and fast protein dynamics, respectively, there is a specific change of the transition energy of every pigment due to the different protein environment [11].

Müh *et al.* [116] present a joint quantum chemical and electrostatic approach to compute the transition energies of pigments in the protein, the so-called site energies, that provides the missing link between crystallography and spectroscopy.

Although, there exist different methods at varying levels of approximation to accurately determine excitonic couplings from structural data, the calculation of site energies has been seen as a challenge because of the complexity of the above interactions [116]. Here, they have found that the calculation of site energies of the FMO complex results in optical spectra that are in quantitative agreement with experimental data.

Different sets of site energies of the seven BChla molecules have been extracted from calculations of optical spectra [11] either by using different fit algorithms or simply by hand. The different calculations predict different pigments to be responsible for the lowest exciton state.

Excitonic couplings, which are a prerequisite for energy transfer, are long-range electrostatic interactions between the local excitations. These couplings cause the excited states to be delocalized - this means that their electronic wave functions contain contributions of several pigments in the complex.

The transfer of population between exciton states depends crucially on excitonic couplings and site energies. Therefore, any clarification on energy-transfer mechanisms based on spectroscopic data and crystal structures requires knowledge of both these quantities which are not directly available from experiments.

Müh *et al.* [116] also revealed a strong influence of the structure of the protein complex on the direction of excitation energy flow within. This result challenged the existing view of how energy flow is regulated in pigment-protein complexes and demonstrated that attention must be paid to the backbone architecture.

Insights into the dynamics of excitons in the FMO complex were obtained from time-resolved non-linear optical spectroscopy which detected sub-picosecond as well as picosecond exciton relaxation times [11]. Two-dimensional photon echo technique [45] permits one to directly detect the coupling between different exciton levels and the energies of exciton states and to infer the spatial and temporal relaxation of excitons from those measures. Still, such a structural interpretation requires information on the site energies and excitonic couplings of the pigments.

In this study, Suzuki *et al.* [131] examine the optimality of site energies and excitonic coupling energies evaluated by various research groups on the basis of generalized quantum master equations describing the EET dynamics among eight BChls. They have found that most energy parameters are nearly optimal for high efficiency in the cases that BChl 1 and/or BChl 6 are initially excited, with some exceptions due to putative inaccuracies in the parameter evaluations.

4 Hierarchical Equations of Motion

One of the viable approaches to explore long-lived quantum coherence and its interplay with the protein environment in EET processes is through the reduced equation of motion. In this approach, the key quantity of interest is the reduced density matrix, i.e., the partial trace of the total density matrix over the environmental degrees of freedom [77]. Typical situations in photosynthetic EET are such that the electronic coupling strengths, between chromophores and their local environment phonons, span a similar range as the reorganization energies, which characterize the time scale of the coupled phonons relaxing to their respective equilibrium states [132, 133]. However, these site-dependent reorganization processes cannot be described by theories that rely on the Markov approximation as it requires the phonons to relax to their equilibrium states instantaneously, that is, the phonons are always in equilibrium even under the electron-phonon interaction.

In order to go beyond the Markov approximation, Tanimura and Kubo [4] developed a new theoretical framework, the HEOM, which can describe the site-dependent reorganization dynamics of environmental phonons. Ishizaki and Fleming [112] adapted this formalism to suit the quantum biological regime which is the form we employ.

The reduced hierarchical equations of motion (HEOM) theoretical framework is a method that can describe the dynamics of a system with a non-perturbative and non-Markovian system–bath interaction at finite temperature, even under strong time-dependent perturbations [4, 82, 83, 134]. In this formalism, the effects of higher-order, non-Markovian system–bath interactions are mapped into the hierarchical elements of the reduced density matrix.

As mentioned in a paper by Chen *et al.* [85], it requires intensive computational techniques to propagate the quantum state of a complete system-environment model explicitly and exactly. Especially when the environment is composed of a huge number of microscopic constituents, an exact description of which is challenging. This demand can be met by the HEOM, however, a common drawback of such a numerically exact solution (the method used does not involve any approximation such that the final result is exact up to a certain level of precision) is its demanding computational resource requirements which can scale poorly depending on the spectral density of the environment, the complexity of the system and the number of independent baths the system is coupled to.

The HEOM approach is numerically exact for the models with the Drude-Lorentz spectral density function employed in this thesis. Research has been done for cases of more general bath models and for dynamics at low temperatures [135–137]. The intermediate system-environment coupling regime

26

is focused on here which is relevant to photosynthetic systems such as the Fenna-Matthews-Olson complex and also, the region of validity where most approximations break down.

HEOM is a numerically exact method which accurately accounts for the reorganization process in which the vibrational coordinates rearrange to their new equilibrium positions upon electronic transition from the ground to the excited potential energy surface. It can describe quantum coherent wavelike motion and incoherent hopping in the same framework and reduces to the conventional Redfield [74, 138, 139] and Förster [140, 141] theories in their respective limits of validity. In this chapter, we highlight the theory required to describe EET dynamics in a photosynthetic complex [142, 143].

The total Hamiltonian is composed of the Hamiltonian of the system, bath and system-bath interaction,

$$\hat{H}_{Tot} = \hat{H}_S + \hat{H}_B + \hat{H}_{SB}. \tag{21}$$

The Hamiltonian of the system refers to the electronic states of a complex containing N pigments,

$$\hat{H}_S = \sum_{j=1}^{N} |j\rangle \epsilon_j \langle j| + \sum_{k \neq j}^{N} |j\rangle J_{jk} \langle k|, \tag{22}$$

where ϵ_j is the excited state energy of the jth site and J_{jk} denotes the electronic coupling between the jth and kth sites.

Here we assume that each pigment is coupled to a separate bath. The bath Hamiltonian represents the environmental phonons,

$$\hat{H}_B = \sum_{j=1}^{N} \hat{H}_{B_j}, \quad \hat{H}_{B_j} = \sum_{\alpha} \hbar \omega_{j,\alpha} \left(\frac{\hat{p}_{j,\alpha}^2 + \hat{q}_{j,\alpha}^2}{2} \right), \tag{23}$$

where $\hat{p}_{j,\alpha}$ is the conjugate momentum, $\hat{q}_{j,\alpha}$ is the dimensionless coordinate and $\omega_{j,\alpha}$ is the frequency of the jth site and αth phonon mode, respectively. The last term of Eq. (21) represents the fluctuations in the site energies caused by the phonon dynamics,

$$\hat{H}_{SB} = \sum_{j=1}^{N} \hat{u}_j |j\rangle \langle j|, \quad \hat{u}_j = \sum_{\alpha} g_{j,\alpha} \hat{q}_{j,\alpha}, \tag{24}$$

where $g_{j,\alpha}$ is the coupling constant between the jth site and αth phonon mode.

The spectral density $\mathcal{J}_j(\omega)$ specifies the coupling of an electronic transition of the jth pigment to the environmental phonons through the reorganization energy λ_j and the timescale of the phonon relaxation γ_j. Here it is expressed as the Ohmic spectral density with Lorentz-Drude cut-off, $\mathcal{J}_j(\omega) = 2\lambda_j\gamma_j\omega/(\omega^2 + \gamma_j^2)$.

We focus on the application of this theory to EET at physiological temperatures of around 300 K. Hence, when the high-temperature condition characterized by $\hbar\gamma_j/k_B T \ll 1$ is imposed, the following hierarchically coupled equations of motion is obtained [112],

$$\frac{\partial}{\partial t}\hat{\sigma}(\mathbf{n},t) = -\left(i\hat{\mathcal{L}}_e + \sum_{j=1}^{N} n_j\gamma_j\right)\hat{\sigma}(\mathbf{n},t) + \sum_{j=1}^{N}\left[\hat{\Phi}_j\hat{\sigma}(\mathbf{n}_{j+},t) + n_j\hat{\Theta}_j\hat{\sigma}(\mathbf{n}_{j-},t)\right]. \qquad (25)$$

In Eq. (25), $\mathbf{n} \equiv (n_1, n_2, \ldots, n_N)$ for sets of non-negative integers and $\mathbf{n}_{j\pm}$ differs from $\mathbf{n}$ by changing the corresponding n_j to $n_j \pm 1$. Furthermore in Eq. (25), the element $\hat{\sigma}(\mathbf{0},t)$ is identical to the reduced density operator $\hat{\rho}(t)$, while the rest are auxiliary density operators. Moreover, the Liouvillian corresponding to the Hamiltonian $\hat{H}_S$ is denoted by $\hat{\mathcal{L}}_e$ and the relaxation operators $\hat{\Phi}_j$ and and $\hat{\Theta}_j$ are given by Eqs. (26), (27) and (28),

$$\hat{\mathcal{L}}_e = [\hat{H}_S, \hat{\rho}_S], \qquad (26)$$

$$\hat{\Phi}_j = iV_j^\times, \quad V_j^\times y = [V_j, y], \qquad (27)$$

$$\hat{\Theta}_j = i\left(\frac{2\lambda_j}{\beta\hbar^2}V_j^\times - i\frac{\lambda_j}{\hbar}\gamma_j V_j^\circ\right), \quad V_j^\circ y = \{V_j, y\}. \qquad (28)$$

Formally the hierarchy in Eq. (25) is infinite and cannot be numerically integrated. In order to make this problem tractable, the hierarchy can be terminated at a certain depth. There are several methods of doing so and in this work we have chosen the following termination condition following Ishizaki and Fleming [142]. For the integers $\mathbf{n} = (n_1, n_2, \ldots, n_N)$ and for characteristic frequency ω_e of $\hat{\mathcal{L}}_e$ where

$$\mathcal{N} \equiv \sum_{j=1}^{N} n_j \gg \frac{\omega_e}{\min(\gamma_1, \gamma_2, \ldots, \gamma_N)}, \qquad (29)$$

Eq. (25) is replaced by

$$\frac{\partial}{\partial t}\hat{\sigma}(\mathbf{n}, t) = -i\hat{\mathcal{L}}_e\hat{\sigma}(\mathbf{n}, t). \tag{30}$$

To elaborate on the truncation, Eq. (31) is the HEOM of the first non-trivial case of a dimer ($N = 2$) where the depth of the hierarchy is $\mathcal{N} = 2$. This case is chosen for its simplicity while it contains all three components of the hierarchy – the density matrix of the system where $n = (0, 0)$,

$$\frac{\partial}{\partial t}\hat{\rho} = -i\hat{\mathcal{L}}_e\hat{\rho}(t) + \hat{\Phi}_1\hat{\sigma}^{(1,0)}(t) + \hat{\Phi}_2\hat{\sigma}^{(0,1)}(t). \tag{31a}$$

the generic HEOM for non-negative integers $(n_1, n_2) = [(0, 1), (1, 0)]$ with auxiliary density operators $\hat{\sigma}^{(n_1,n_2)}(t)$,

$$\frac{\partial}{\partial t}\hat{\sigma}^{(n_1,n_2)}(t) = -\left(i\hat{\mathcal{L}}_e + n_1\gamma_1 + n_2\gamma_2\right)\hat{\sigma}^{(n_1,n_2)}(t) + \hat{\Phi}_1\hat{\sigma}^{(n_1+1,n_2)}(t) + n_1\hat{\Theta}_1\hat{\sigma}^{(n_1-1,n_2)}(t)$$

$$+ \hat{\Phi}_2\hat{\sigma}^{(n_1,n_2+1)}(t) + n_2\hat{\Theta}_2\hat{\sigma}^{(n_1,n_2-1)}(t) \tag{31b}$$

and terminating term of the HEOM sequence for $(n_1, n_2) = [(0, 2), (2, 0), (1, 1)]$,

$$\frac{\partial}{\partial t}\hat{\sigma}^{(n_1,n_2)}(t) = -i\hat{\mathcal{L}}_e\hat{\sigma}^{(n_1,n_2)}(t). \tag{31c}$$

The reduced hierarchy equation Eq. (25) is employed in the Sections 6 and 7 as it is solved to generate the data sets used in training and testing the machine learning models.

5 Machine Learning

Machine learning has become a tool to accelerate resource demanding computations and experiments in many fields, particularly the physical sciences [56]. Although it is a sub-discipline of artificial intelligence, other important parent disciplines of machine learning are statistics, computer science and mathematics. This field, of data-driven decision making or prediction, largely consists of algorithmic systems and statistical models capable of performing defined tasks without being provided specific instructions. Furthermore, machine learning models can be seen as sets of equations, rules or algorithms that infer information relevant to a task by identifying and exploiting statistical correlations from the data ensembles provided, and consequently learn how to solve these tasks.

While data driven regression methodologies are more conventional, developments in machine learning led to significant breakthroughs in material and drug design [144]. This motivated further research for the advancement of light-harvesting applications, particularly for organic light-emitting diodes [145] and photo functional molecules [146].

To be able to analyse the behaviour of light-harvesting molecules, precise modelling of these systems over long time scales is required. However, computational modelling of large, complex systems poses extreme challenges to standard simulation methods. Simulation methods enhanced with machine learning have the potential to overcome this caveat. In the context of biophysical sciences, the first molecular dynamics simulations with a purely machine learning based ground state density functional have been reported [57].

Machine learning models are able to produce different types of outputs, targets, labels or predictions. As when future values of a time series have to be predicted, it is a forecast and when the content of images is recognised, it is a classification.

5.1 Constructing a model

This section provides a broad overview of the model construction process. The categories of machine learning and related processes are discussed in detail in Chapters 6 and 7.

One commonly employed category within machine learning is supervised learning. Tasks in this category are characterised by models that are trained to predict a set of outputs (labels) based on a set of inputs (features). Hence, the models must learn a mapping which assigns given features to their associated labels.

On the other hand, the category of unsupervised learning approaches does not aim to directly make predictions based on features. The focus in these cases is on deducing the a priori probability distribution. Therefore, unsupervised learning has the potential to reveal patterns in the provided dataset. A third area of machine learning is reinforcement learning, in which an agent gets rewarded or punished for certain decisions according to a given rule, and the agent learns an optimal strategy by trial and error. In this thesis, I will focus on supervised learning problems only.

The basic structure of a supervised learning task is as follows: Given an input domain $\mathcal{X}$ and an output domain $\mathcal{Y}$, a training data set $\mathcal{D} = (x^{(1)}, y^{(1)}), ..., (x^{(M)}, y^{(M)})$ of M training pairs $(x^{(M)}, y^{(M)}) \in \mathcal{X} \times \mathcal{Y}$, as well as a new unclassified input $\tilde{x} \in \mathcal{X}$, the task is to guess or predict the corresponding output $\tilde{y} \in \mathcal{Y}$.

The data is pre-processed in a crucial step before the model is constructed [147]. This involves feature extraction to compress the raw data and feature scaling such as shifting the data to a zero mean and unit variance which helps to avoid the impact of widely different scales. The output range determines the kind of supervised problem to be solved as when $\mathcal{Y}$ is a set of $\mathcal{D}$ discrete class labels $c_1, ..., c_D$ one speaks of a classification task and regression tasks require $\mathcal{Y}$ to reside in an interval of the real numbers.

A high-level category of model family is chosen once the data has been prepared to limit the kind of possible functions that may be used. During the training process, the model is presented with samples of pairs to extrapolate the underlying structure-property relation. As the model is optimized through training, it leverages correlations from the data as opposed to relying on the laws of physics. The model is adapted to the data such that a specific model function is chosen from the broader category. It is important to mention that machine learning models for supervised tasks can only identify dependencies of the properties on variables that are included in the dataset. For example, a temperature dependence will not be revealed if temperature is not provided as one of the factors in the feature dataset. Once the model is adapted to the data it may be used for prediction.

The model training procedure translates to estimating the best set of model parameters given the data. Training typically means defining an objective function that quantifies the quality of a model, hence, the greater problem transforms into optimisation problems that requires powerful computational resources [148]. The dataset is divided into a training set and a test set as a portion of the data is used for evaluation of the model performance. While the model is built based on the training set, its reliability is measured by how accurately it performs based on unseen ensembles found in the test set. The error on the training set is termed the empirical error and the error on the test set is called the generalisation error. A significant point is that a low empirical error does not necessarily lead to a

low generalisation error. A loss function is introduced to measure the errors. The training process minimizes the loss function, therefore, solving an optimization problem. A common choice for the loss function is the mean squared error (MSE) described in Eq. 32 where y_i and f_i represent the ith value to be predicted and the predicted value of y_i, respectively. This equation compares the outputs produced by the model when given the inputs to the ground truth i.e. the target outputs.

$$MSE = \sum_{i=1}^{N} \frac{(y_i - f_i)^2}{N}. \tag{32}$$

There exist various schemes for solving optimization problems in the machine learning regime to find a global solution. Popular methods include iterative searches such as gradient descent.

In gradient descent methods, the algorithm iteratively calculates the next point using the gradient at the current position then scales it by a learning rate η and subtracts the obtained value from the current position i.e. makes a step. The value is subtracted to minimize the function. The parameters θ of an objective function $f(\theta)$ are successively updated according to

$$\theta^{(t+1)} = \theta^{(t)} - \eta \nabla f(\theta^{(t)}). \tag{33}$$

The advantage lies in its simplicity and applicability in many settings. An important variation of gradient descent is stochastic gradient descent where smaller alternating batches of the training set are evaluated in the objective function. Another variation stems from a step-dependent learning rate $\eta(t)$.

Desired models exhibit a balance between explaining the training dataset sufficiently well without overfitting which occurs when the model learns the particulars of the training set instead of reflecting the general structure of the physical law behind the data. For example, a high-order polynomial can satisfy every data point and fit the training data perfectly well, however, the high-order polynomial has large errors when considering the test set. While training normally continuously reduces the empirical error, the error on the test set may increase after a certain number of steps in the training procedure. This is an indicator of overfitting. Regularisation constraints are used to avoid overfitting and can be as simple as early stopping of the training.

The performance of a machine learning algorithm is therefore measured in its ability to generalize.

5.2 Artificial neural networks

Deep learning is a sub-field of machine learning. Neural networks, also known as artificial neural networks (ANNs), make up the backbone of deep learning algorithms (see Figure 3). These models

were inspired by biological neural networks [149, 150] and they have a graphical representation reminiscent of neurons that are connected by synapses in the brain as depicted in Figure 4.
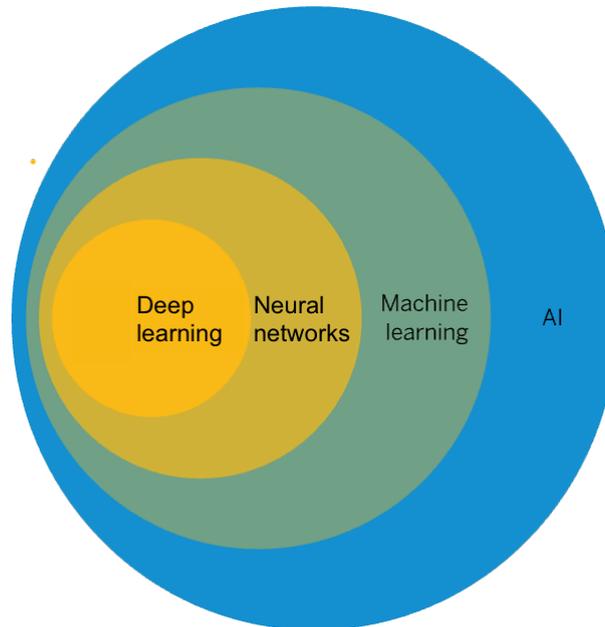
Fig. 3: Diagram of the relationship between AI, deep learning, machine learning and neural networks.

Research centred around neural networks has advanced over the last few decades, however, the history of neural networks is longer than most people think. In 1943, Warren S. McCulloch and Walter Pitts published "A logical calculus of the ideas immanent in nervous activity" which sought to understand how the human brain could produce complex patterns through connected brain cells or neurons [149]. One of the main ideas that came out of this work was the comparison of neurons with a binary threshold to Boolean logic. In 1958, Frank Rosenblatt developed the perceptron [151]. He took McCulloch and Pitt's work a step further by introducing weights to the equation. Rosenblatt demonstrated that a computer could learn how to distinguish cards marked on the left vs. cards marked on the right. The rediscovery of the backpropagation algorithm occured in the late 80s [152, 153], although it was initially developed by Paul Werbos in 1974 [154]. In 1982, Hopfield depicted that a certain type of network recovers properties of associative memory [150]. In 1989, LeCun published a paper [155] illustrating how the use of constraints in backpropagation and its integration into the neural network architecture can be used to train algorithms and recent developments in 'deep' neural network architectures outperforming several benchmarks in machine learning research are given by Hinton [156].

A neural network comprises four main components: inputs, weights, a bias and an output. A basic example of a network is constructed in the following fashion: input layer, hidden layers and output layers. For each hidden layer of the network, the bias and weights are used in an algebraic formula
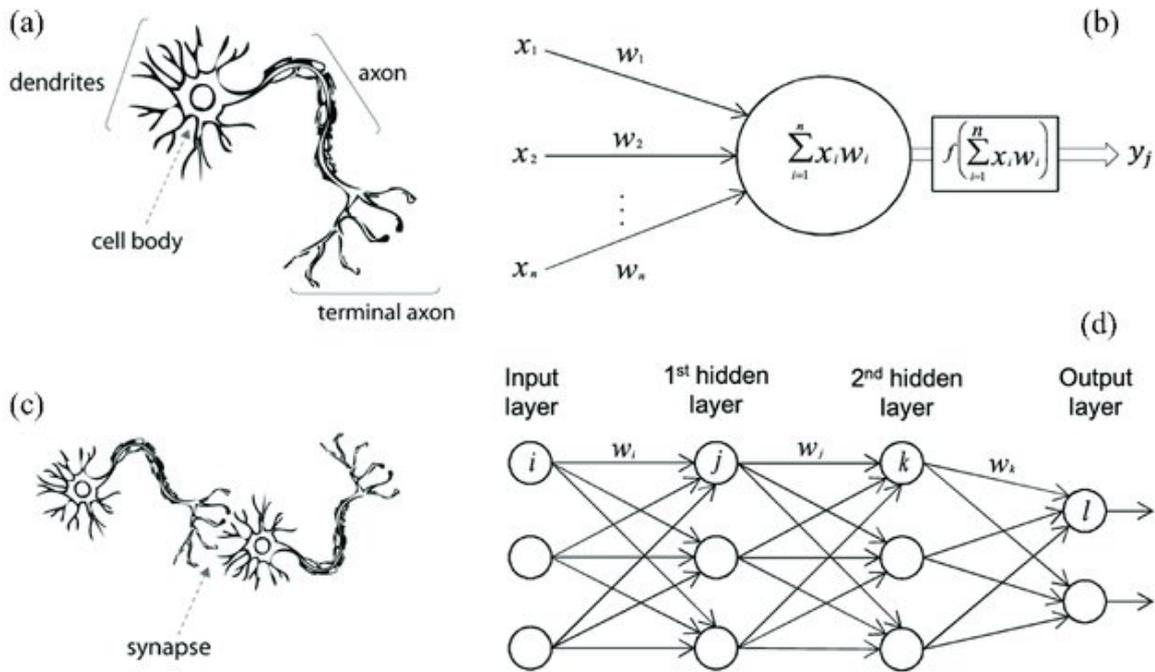
Fig. 4: A biological neuron in comparison to an artificial neural network: (a) human neuron, (b) artificial neuron, (c) biological synapse and (d) ANN synapses [157].

that quantify the importance of the varying inputs to the outputs. An activation function is applied on each layer to determine which nodes (neurons) in the layer will be permitted to pass information onto the next layer. The activation function in this set-up corresponds to the 'firing' of a biological neuron [158]. Once all the outputs from the hidden layers are generated, then they are used as inputs to calculate the final output of the neural network. Neural networks rely on training data to learn and improve their accuracy over time. What makes neural networks different from regression is that in regression, a weight can be changed without affecting the other inputs in a function. In the case of neural networks, the output of one layer is passed into the next layer of the network, hence, a single change can have a cascading effect on the other neurons in the network.

The "deep" in deep learning refers to the number of layers in a neural network. One that consists of more than three layers—inclusive of the inputs and the output—can be considered a deep learning algorithm. Classical or "non-deep" machine learning is more dependent on human intervention to learn compared to deep learning which automates much of the feature extraction piece of the process, eliminating some of the manual human intervention required.

As a model is trained, its accuracy is evaluated using a cost (or loss) function. A neural network model is usually trained through backpropagation - short form for "backward propagation of errors". This method helps calculate the gradient of a loss function with respect to all the weights in the network.

Backpropagation allows us to calculate and attribute the error associated with each neuron, allowing us to adjust and fit the parameters of the model correctly. It is the method of fine-tuning the weights of a neural network based on the empirical error obtained in the previous iteration of the training procedure. Adjusting of the weights allows the error rates to reduce, in turn making the model reliable by increasing its generalization.

Perceptrons are the basic building block of artificial neural networks, created by Frank Rosenblatt in 1958, with only a single neuron it is the simplest neural network.

As opposed to being organised in layers, the graphical representation of recurrent neural networks (RNNs) is an all-to-all connected graph of units or feedback loops which are collectively updated in discrete time steps. Information is therefore not fed forward through layers, rather it is propagated in time. The input can therefore be understood as the state that some units are set to at time $t = 0$, while the output can be read from one or more designated units at a later time T. The units that are neither fed with input nor used to read out the output are called 'hidden units' and serve computational purpose. These learning algorithms are primarily leveraged when using time-series data to make predictions about future outcomes.

6   Machine learning for excitation energy transfer dynamics

*This chapter is based on Ref [159]: Kimara Naicker, Ilya Sinayskiy, and Francesco Petruccione Physical Review Research **4**, 033175 (2022).*

A well-known approach to describe the dynamics of an open quantum system is to compute the master equation evolving the reduced density matrix of the system. This approach plays an important role in describing excitation transfer through photosynthetic light harvesting complexes (LHCs). The hierarchical equations of motion (HEOM) was adapted by Ishizaki and Fleming (J. Chem. Phys., 2009) to simulate open quantum dynamics in the biological regime. We generate a set of time dependent observables that depict the coherent propagation of electronic excitations through the LHCs by solving the HEOM. The computationally intractable problem here is addressed using classical machine learning (ML). The ML architecture constructed here is of model character and it is used to solve the inverse problem for open quantum systems within the HEOM approach. The objective is to determine whether a trained ML model can perform Hamiltonian tomography by using the time dependence of the observables as inputs. We demonstrate the capability of convolutional neural networks to tackle this research problem. The models developed here can predict Hamiltonian parameters such as excited state energies and inter-site couplings of a system up to 99.28% accuracy.

6.1   Introduction

During the first step of photosynthesis in light harvesting complexes, photons are absorbed by the antenna. While part of the energy of the photons is converted into heat in the form of molecular vibrations, most of the energy is captured as excitons which are subsequently transferred via chromophores to the reaction centre through a process labeled as excitation energy transfer (EET). It is at the reaction centre where photochemical reactions are triggered [20].

The evidence for quantum coherence, which has no classical analogue, in the exciton transport process became undeniable in 2007 when Engel *et al.* [1] used two-dimensional spectroscopic signatures [160] to demonstrate quantum 'beating' within a photosynthetic complex at 77 K, a result that was later confirmed at room temperature. Quantum beating in spectroscopic measurements provides a direct measure of quantum coherence on the appropriate energy and time scales [161]. While electronic coherence was first proposed as the source for the observed long-lived quantum coherence [1], experimental and theoretical evidence has also supported proposals that the phenomenon resulted from a mixture of electronic and vibrational states which is referred to as

'vibronic' coherence [162].

The idea of quantum coherence playing a role in photosynthesis arose from observations that some energy or electron transfer processes in bacterial and plant pigment–protein complexes are efficient to an extent that exceeds explanation using only classical theory. Engel *et al.* [1] investigated photosynthetic EET in the Fenna–Matthews–Olson (FMO) protein of green sulfur bacteria [20, 115]. The FMO complex was the first chlorophyll-containing protein that was crystallized [24]. Prior to mass spectrometry measurements of the protein that confirmed the existence of an eighth pigment, it was accepted that the protein consisted of seven pigments [163]. Subsequently, FMO is made up of three subunits each consisting of eight bacteriochlorophyll molecules. The FMO complex serves as a bridging energy wire as it is tasked with transporting energy in the form of light harvested in the antenna chlorosome to the reaction centre pigments [20]. It represents an important model in EET and has been extensively studied experimentally and theoretically. Engel and collaborators succeeded in observing long-lasting quantum effects providing direct evidence for long-lived electronic coherence [161]. The observed coherence lasts for time scales similar to the EET timescales, implying that electronic excitations move coherently through the FMO protein rather than by previously proposed incoherent hopping motion [164, 165]. Panitchayangkoon *et al.* [3] presented evidence that quantum coherence survives in FMO at physiological temperature for at least 300 fs which is long enough to impact biological energy transport. Collini *et al.* [44] made observations that provide evidence for quantum coherent sharing of electronic excitation across proteins under biologically relevant conditions. They suggest that distant molecules within the photosynthetic proteins are 'wired' together by quantum coherence for more efficient light-harvesting in cryptophyte marine algae. Lee *et al.* [120] present experimental results in which they suggest that correlated protein environments preserve electronic coherence in photosynthetic complexes and allow the excitation to move coherently in space which enables highly efficient energy harvesting and trapping in photosynthesis. However, explanations for observed long-lived coherence have evolved during the past decade [166–168]. More recently, Fuller *et al.* [162] and Thyrhaug *et al.* [169] suggest that the coherences are of a mixed electronic-vibrational nature and may enhance the rate of charge separation in oxygenic photosynthesis.

Understanding the relationship between the structure of light harvesting complexes and their excitation energy transfer dynamics is of importance in many applications. Insight into long-lived quantum coherence in EET processes can be gained through the reduced equation of motion and the numerically exact formalism of quantum dynamics adopted to study EET processes is the hierarchical equations of motion (HEOM) derived by Tanimura and Kubo [4].

Machine learning is a well established tool that has been actively applied in various ways to address physical problems [7]. One common strategy is to use supervised learning in which an algorithm is trained with datasets that are labeled beforehand then, the goal of the algorithm is to establish a general rule for assigning labels to data outside the training set. This approach can be used to identify distinct phases of matter and the transitions between them, one of the central problems in condensed-matter physics, that has been tackled by Carrasquilla and Melko [170]. Machine learning techniques have been used to represent and solve quantum systems such as in a work by Carleo and Troyer [171] where the authors introduce an ansatz capable of both finding the ground state and describing the unitary time evolution of complex interacting quantum systems. More specifically and in a bid to investigate open quantum systems further, the applications of supervised machine learning that we are interested in are related to forms of approximating solutions to open quantum system dynamics such as where multi-layer perceptrons have been used to obtain the exciton dynamics of large photosynthetic complexes [22] and to better understand the relationship between the structure of light-harvesting systems and their excitation energy transfer properties [172]; where recurrent neural networks were used to model quantum systems interacting with an unknown environment [132] and where convolutional neural networks were used to predict long-time dynamics of an open quantum system [65].

More specifically, research has been conducted where varying techniques have been employed to address inverse problems in open quantum systems. Vargas-Hernández *et al.* probe the steady state (SS) solution of the Liouvillian in relation to the computation of desired physical observables [173]. They present a novel methodology to address the inverse design of quantum systems interacting with multiple environments. This methodology, based on automatic differentiation, is capable of differentiating the SS solution with respect to any parameter of the Liouvillian. The authors claim that their approach has a low memory cost and is agnostic to the exact algorithm for computing the SS. An advantage of this method is illustrated in the text by inverse designing the parameters of a quantum heat transfer device that maximizes the heat current and the rectification coefficient. Hou *et al.* investigate Hamiltonian tomography through a machine learning approach [174]. Particle or energy transfer through quantum networks is determined here by network topology and couplings to environments. They study the combined effect of these characteristics on the efficiency of quantum transfer through quantum networks. They are able to determine the Hamiltonian parameters corresponding to minimum transfer time by Bayesian optimization. Their approach can be applied to determine quantum speed limits in other applications. In another study, Vargas-Hernández *et al.* further prove that for the common task of the optimization of physical models to reproduce a set of target observables, machine-learning algorithms are useful [175]. In the case where multiple target observables are considered, the authors used a multi-objective optimization protocol where the goal

is to learn the limits of each objective through the Pareto front [176]. They illustrate that more can be learned about the robustness of a physical model by finding the Pareto front than can be done with results from using a single-objective optimization scheme. A significant biophysical process retinal photoisomerization, as it occurs in nature, i.e., in the steady state induced by incoherent radiation, is a particularly important system to examine via this algorithm.

The primary focus of this work is on using classical machine learning models to study the quantum dynamics of EET. We can generate a time dependent set of observables that depict the coherent movement of electronic excitations through a photosynthetic pigment-protein complex by solving the HEOM. Here we develop a scalable and efficient tool for the description of the dynamical properties of open quantum systems by use of a trained convolutional neural network (CNN) to solve the inverse problem. This means that the objective is to determine whether a trained CNN can accurately describe the system under study, by predicting the parameters of the system Hamiltonian such as excited state energies and inter-site couplings, when given this time dependent data of varying length (see Figure 5).

The HEOM framework used in this chapter is described in Chapter 4.

There exist numerous studies [22, 65, 132, 172, 177, 178] of machine learning techniques applied to accelerate computations by many orders of magnitude at a reasonable level of accuracy. A machine learning model can be leveraged to predict the reduced density matrix of a system given the Hamiltonian of the system. In this approach, the model is trained and tested on a large and diverse enough dataset of Hamiltonians and corresponding reduced density matrices such that it may learn patterns in the data and be able to present highly accurate predictions without having any knowledge of the theory or in this case, HEOM.

However, in an experimental setting one may gather certain time-dependent observational data and subsequently need to use these findings to attain the Hamiltonian of the excitonic system under study. The use of machine learning models in this work is to act as a *blackbox* which one can input excitation energy transfer observations into and obtain Hamiltonian parameters from.

In the subsequent sections, Section 6.2 we describe the machine learning basics required to follow the study. Thereafter, in Section 6.3 we generate multiple datasets comprising excited state population dynamics and corresponding Hamiltonian parameters and in Section 6.4 we design the supervised machine learning model architecture to be used for making predictions based on the generated datasets.
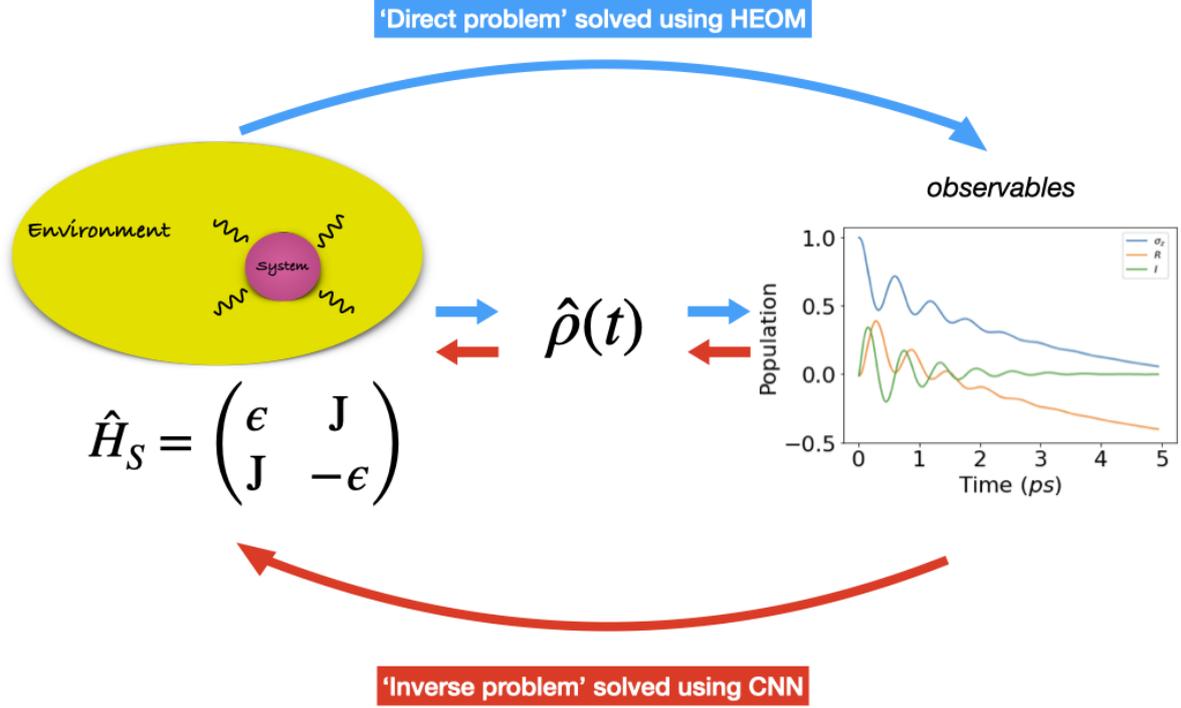
Fig. 5: Machine learning Hamiltonian tomography in open quantum systems. The 'direct problem' is the process of using the Hamiltonian of an open quantum system $\hat{H}_S$ as an input to the HEOM to calculate the evolution of the reduced density matrix $\hat{\rho}(t)$ to produce the time dependent observables. The 'inverse problem' addressed in this study is the process of using the observables as an input to the CNN to reproduce the system Hamiltonian.

6.2 Supervised machine learning

Machine learning algorithms are used to learn underlying patterns embedded in the data. In the realm of classical machine learning, there exist three broad types classified by the amount and type of supervision models get during training: supervised, unsupervised and reinforcement learning. In supervised learning, the training set fed to the model includes the true solutions called labels [179, 180].

A neural network is a machine learning model whose structure is inspired by the networks of biological neurons found in our brains [149]. They are made up of layers of neurons which are core processing units of the network. Usually these layers contain an input layer to receive input features, an output layer to make final predictions and hidden layers which perform most of the computations done by the network. Convolutional neural networks (CNNs) [181] is a class of neural networks which emerged from the study of the brain's visual cortex [182, 183]. CNNs specialize in processing data that has

a grid-like topology. The human brain processes information when we see an image as each neuron works in its own restricted region of the visual field called the receptive field and is connected to other neurons in a way that covers the entire visual field. Just as each neuron responds to stimuli only in its receptive field in the biological vision system, each neuron in a CNN processes data only in its receptive field. The layers are arranged such that they detect simpler patterns first and more complex patterns deeper into the network. A rich description of how the convolution operation works and the advantages of using CNNs has been written by Goodfellow *et al.* [184].

In this work, machine learning has been leveraged to solve the HEOM *inversely* without explicitly solving the equations at all so that predictions of the parameters of the Hamiltonian of a system can be made when given time dependent observations.

6.3 Generating the database

To demonstrate the capabilities of our machine learning approaches for the regression task at hand, we investigate three datasets of increasing complexity that are randomly generated excitonic systems. The parameters of the Hamiltonians in these datasets are motivated by and sampled around the same order of magnitude as those that are typical of the light-harvesting pigment-protein FMO (see Table 1).

We impose a linear chain such that only neighbour-neighbour couplings are permitted and for simplicity, we consider that the transition rates are strictly real in value. When sampling the Hamiltonian parameters we consider the excited state energy ϵ_j and inter-site coupling J_{jk} with respect to ϵ_1. Hence, the Hamiltonian of an N-pigment system would require $2(N-1)$ real parameters. In constructing the first dataset, we consider two-level excitonic systems which allow for excitation energy transfer between two excited states. In this case, there are two Hamiltonian parameters needed to describe the two-level system as seen in Eq. (22) where $N = 2$. Therefore, by solving the HEOM we obtain the time evolution of an N-dimensional reduced density matrix for each sample. In a similar way, the second and third datasets consider three-level and four-level systems which require four and six Hamiltonian parameters, respectively.

The following data is captured and stored for each sample in each dataset: Hamiltonian parameters which are used as labels for the machine learning model and, the time evolution of each element of the reduced density matrix to be used as input features for the machine learning model. An example of an input feature for a two-level system can be seen in Figure 6 which depicts the time evolution of the population of site 1 which corresponds to the first element in the reduced density matrix. For each dataset, 25 000 Hamiltonians were generated by sampling a uniform distribution for excited state energies and inter-site couplings within a fixed range of values around those that are typical of FMO

complex shown in Table 1. Only neighbour-neighbour couplings were considered to be non-zero, hence, J_{jk} were only sampled for each site's nearest neighbours. Furthermore, the values of excited state energy ϵ_j have been sampled with respect to 12 400 cm^{-1} for all sites [108, 172].

Table 1: Lower and upper limits in between which excited state energies ϵ_j and inter-site couplings J_{jk} for each site's nearest neighbours were uniformly sampled to generate the three datasets of this study.

Dataset	$\epsilon_j[cm^{-1}]$	$J_{jk}[cm^{-1}]$
2	[-100, 100]	[-100, 100]
3	[-100, 100]	[-100, 100]
4	[-100, 100]	[-100, 100]
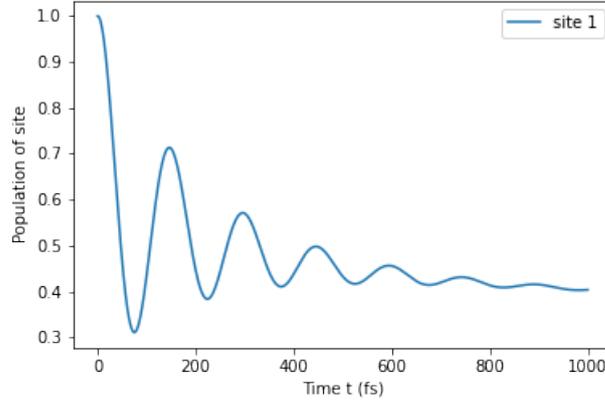
Fig. 6: Time evolution of the population of site 1 of a 2-level system calculated by the HEOM, Eq. (25). This calculation was done at T = 300 K where the reorganization energy and the phonon relaxation time are set to be $\lambda = 35$ cm^{-1} and $\gamma = 106.1767$ cm^{-1}, respectively. Other parameters were fixed to be E = 100 cm^{-1} and $J_{12} = 100$ cm^{-1}.

By explicitly solving the HEOM, we compute the time evolution of the reduced density matrix for each Hamiltonian in each of our datasets. As opposed to the traditional 4th-order Runge-Kutta method, the numerical propagation was done by an exponential expansion method described by Dattani *et al.* for 1 ps represented by 5 000 time steps [167]. Each computation took approximately 1 s, 1.5 s and 2 s for a two-level, three-level and four-level system sample, respectively. For all samples in all datasets, the HEOM were truncated at a depth of 3 (refer to Eq. (29)). For all Hamiltonians we assumed identical Drude–Lorentz spectral densities describing the influence of the bath on each excited state. For all datasets, simulations were done at temperature T = 300 K, reorganization energy $\lambda = 35$ cm^{-1} and phonon relaxation time $\gamma = 106.1767$ cm^{-1} [112]. Python's main scientific libraries used are NumPy, pandas, and Matplotlib and Python frameworks for machine learning tasks are Scikit-Learn, TensorFlow and Keras.

6.4   Model architecture

The architecture of our CNNs are designed for supervised learning of excitation energy transfer dynamics. As mentioned in Subsection 6.3, the elements of the reduced density matrices obtained by solving the HEOM are the inputs known as features for the CNN. To reduce the computational cost of model training and to improve the performance of the model, a feature selection process was carried out to reduce the number of input variables. During the process of optimization of our CNNs it was found that rather than the entire reduced density matrix, only the diagonal elements and their nearest-neighbours were required as input features to allow the models to perform best. Data scaling was performed to transform all datasets as a pre-processing step by fitting a scaling object only to the training data then using it to tranform the training/ validation and test sets. All features of all datasets were normalized by rescaling the data into the range [0,1]. Similarly, the labels were transformed so as to be normally distributed such that the mean of the values is 0 and the standard deviation is 1. For each dataset, all features were reshaped into 4-dimensional tensors (number of samples, number of time steps, number of features, 1) and provided as input features to the CNNs, which were used to predict excited state energies and inter-site couplings known as labels. Since the input features of neural networks need to be of fixed size, we construct separate CNNs for each dataset in order to treat the different dimensionalities of the Hamiltonians. These CNNs only differ by their input and output shapes. The hidden layers of the CNNs consist of two 2D convolutional layers with by a max pooling operation layer, the previous three layers are repeated and then followed by a flattening layer and three dense layers. Goodfellow *et al.* [184] describe general guidelines for choosing which architectures to use in which circumstances. The full architecture of the models with hyperparameters can be seen in Figure 13 in the appendix.

The 25 000 Hamiltonians of each dataset were split into three sets: a training set of 85% of all Hamiltonians where 80% of these are used for training CNN model instances with particular hyperparameters and the other 20% form a validation set used to evaluate the CNN architecture during optimization of the hyperparameters and a test set of 15% of all Hamiltonians to probe out-of-sample prediction accuracies. Noteworthy, after splitting the data into train, validation and test sets, the distribution of the Hamiltonian parameters (labels) of each of the subsets maintained their uniform distributions. All constructed CNN models were trained with 100 data points per batch and the ADAM optimizer with a learning rate of 0.001 until the mean squared error (MSE) on the validation set increased over three full consecutive training epochs on a computational cluster.

The MSE provides a direct quantitative check of the extent to which the predicted response value for a given observation is close to the true response value for that observation. This measure works

well in ensuring that our trained model has no outlier predictions with large errors since it allocates larger weight to these errors due to the squaring part of the function. An ideal MSE value is 0.0, which implies that all predicted values matched the expected values exactly. Although the MSE was used to evaluate the cost function throughout the training and testing processes, the coefficient of determination given in Eq. (34) was calculated as an alternate accuracy measure where y_i, f_i and $\overline{y}$ represent the ith value to be predicted, the predicted value of y_i and the mean value of all y_i samples, respectively,

$$R^2 = 1 - \frac{\sum_i (y_i - \overline{y})^2}{\sum_i (y_i - f_i)^2}. \tag{34}$$

The co-efficient of determination R^2 depends on the ratio of total deviation of results described by the model. The reported values in Section 6.5 are presented as a percentage such that $0 \leq R^2 \leq 100$. R^2 can be interpreted as the percentage of variation in the dependent output attribute that the model is capable of explaining. Quantitatively, the closer R^2 is to the upper limit of 100, the better. Both of these measures were used in analyzing the results in order to determine the predictive power of the model. All CNN models were generated and trained using the Tensorflow package.

6.5 Results

In this section, we demonstrate the capabilities of our trained CNN models by analyzing the MSE between predicted Hamiltonian parameters and those used in the numerically exact HEOM calculations and, the coefficient of determination (R^2 score). Our trained models predict Hamiltonian parameters for test (out-of-sample) data at almost the same accuracy as for training and validation data on which CNN model parameters and hyperparameters were optimized. This demonstrates the ability of our models to generalize well to previously unseen data and to provide out-of-sample predictions with high accuracy. To support this conclusion, we present the results of the 5-fold cross-validation of each of our models in Table 2 which was carried out prior to training. A complete set of samples is randomly shuffled and split into the specified number of folds to form smaller sample groups. Cross-validation is a re-sampling procedure then used to evaluate the performance of a machine learning ansatz on each of the limited data sample sets formed. The model is then fitted using the K-1 folds and validated using the remaining Kth fold. This process is repeated until every K-fold has served as a test set then the average of the recorded scores are captured. The reported values in Table 2 are the average MSE and R^2 score values.

We highlight that the K-fold cross validation results are based on the training sub-dataset split into 5 smaller datasets in terms of the number of data samples which further implies less variance amongst

Table 2: K-fold cross-validation results for the CNN model used for each dataset where K=5. The average MSE and R^2 scores are given with their standard deviations.

Dataset	MSE	R^2 score
2	0.0121 ± 0.0001	99.944 ± 0.001
3	0.0557 ± 0.0001	94.734 ± 0.001
4	0.1162 ± 0.0001	91.632 ± 0.002

the data samples. This justifies the observation of much lower average MSE and higher average R^2 score values in Table 2 in comparison to Table 3 where the full cohort of the training dataset samples were processed. As R^2 can be interpreted as the percentage of variation in the dependent output attribute that the model is capable of explaining, the consistency of the R^2 score values per dataset (i.e. 2, 3 or 4) across Table 2 and Table 3 show the reliability of the models developed. Table 3 summarizes the results for the predicted Hamiltonian parameters for our three generated datasets where the full length of 1 ps equating to 5000 timesteps was considered for all selected input features. Furthermore,

Table 3: Mean squared error (MSE) and coefficient of determination (R^2 score) of Hamiltonian parameters used in HEOM calculations and predicted by the trained CNNs. For all three datasets, the full time length of 1 ps for all features were input to the model. The results of the training, validation and test sets are shown, separately.

Dataset	Train		Validation		Test	
	MSE	R^2 score	MSE	R^2 score	MSE	R^2 score
2	0.83	99.16	0.65	99.34	0.70	99.28
3	2.92	97.06	2.86	97.11	3.28	96.64
4	6.58	93.42	6.91	93.04	7.31	92.63

we can deduce that the architectures of the neural networks are well-balanced and neither in the regime of over- or under- fitting which would result in a large discrepancy in errors between the training and validation datasets. The predictions carried out with the CNN architectures only show variation in their performances depending on the dataset under study. Overall we find a high accuracy of our predictions and small mean squared errors on the datasets which are in the range between 0.70 for a 2-level system and 7.25 for the largest considered 4-level system. The 4-level system dataset exhibits the most diverse transfer properties which explains the larger mean squared errors in the predictions when compared to the other datasets.

Figure 7 provides a visual comparison of Hamiltonian parameters as computed with the HEOM approach and predictions with trained CNN models. In Figure 7(a), one can observe that there exist a

group of points of excited state energies which are not well predicted by the model. As random Hamiltonians were generated in the data preparation step, the Hamiltonians that correspond to these points cover the full range of allowed excited state energies as seen in Figure 7(a), however, all of these points also correspond to the range of inter-site couplings that is [-30, 30] cm^{-1}. This is known as the overdamped regime where the time dependent observables do not display a coherent but rather a purely dissipative behaviour, hence, the CNN does not perform well in differentiating between these represented systems.

As highlighted in Section 6.3, feature selection is critical to the performance of trained models. The time evolution of selected elements of the reduced density matrices used as input features each contain data for 1 ps long. The next aim of this work was to determine by how much this time series data could be shortened by while still maintaining the accuracy measures achieved with the full dataset. The results obtained are shown in Table 4.

Table 4: Mean squared error (MSE) and coefficient of determination (R^2 score) of Hamiltonian parameters used in HEOM calculations and predicted by the trained CNNs. For each dataset, the time length in fs for all features which were input to the model are given. The results of the training, validation and test sets are shown, separately.

Dataset	Time	Train		Validation		Test	
		MSE	R^2 score	MSE	R^2 score	MSE	R^2 score
2	400	0.77	99.22	0.95	99.04	0.96	99.03
3	400	2.39	97.59	3.01	96.95	3.02	96.91
4	400	6.62	93.36	7.82	92.08	7.27	92.70

From this we can deduce that the full time length is not required to maintain high accuracy, rather only 400 fs are sufficient for 2, 3 and 4 level systems. The observed prediction errors are also consistent with the complexity of the system dynamics for each of the three datasets which indicates that CNN models generally benefit from a wider sampling of the input parameter space.

6.6 Conclusions

During photosynthesis in light harvesting complexes, energy is transferred from antenna pigments to the reaction center to trigger photochemical reactions. The formalism adopted to study excitation energy transfer (EET) processes is the hierarchical equations of motion (HEOM). This work focuses on leveraging classical machine learning models to study the dynamics of EET within open quantum systems. We propose the use of a trained convolutional neural network (CNN) to perform Hamiltonian tomography when given input data that represents the dynamics of EET through the open quantum
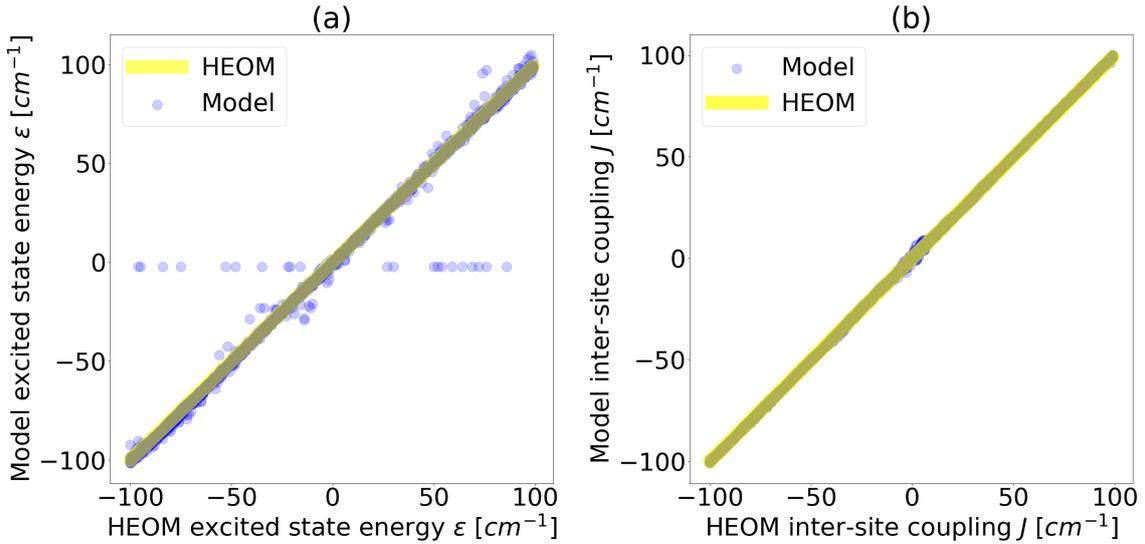
Fig. 7: (a) Excited electronic energy and (b) electronic coupling as computed with the HEOM approach compared to prediction from CNN model for a two-level system. The blue dots represent the model predictions. The yellow line indicates perfect agreement between HEOM results and predictions by the model.

system over time. We have discussed the investigation of EET for 2-, 3- and 4-level systems where linear chain configurations were imposed. The performance of the models were gauged by mean-squared error and coefficient of determination measures. We have proven the capabilities of CNNs and supervised machine learning as an efficient tool for solving the inverse problem of the HEOM by employing a model to predict the parameters of Hamiltonian's when given underlying time dependent observations as features. In particular, we have shown that using a trained CNN one can predict the Hamiltonian parameters such as excited state energy and electronic coupling up to 99.28% accuracy and mean-squared error as low as 0.65. We propose the use of the trained CNNs as an efficient way to study the excitation energy transfer dynamics of biological complexes. An improvement that can be investigated in future is a more sophisticated algorithm that will be able to distinguish between the systems in the overdamped regime that can be described by the set of observables that are purely dissipative. This work can be developed further in a few ways, either by investigating fully connected neural networks or by developing a model such that it may be independent of the dimension of input data. The latter modification would allow the user to input data such that in return the model may determine the dimension of the system under study which represents the number of pigments in the complex. The linear chain model imposed here permits only nearest-neighbour coupling. Further investigations regarding the effects of additional stronger (up to 1000 cm^{-1} [185]) couplings would better represent complex systems as well as further development of the model such that it is capable of high accuracy results when processing data based on greater N-pigment systems. This study serves

as a proof-of-principle piece. Although it involves a rather crude approximation for a real system like FMO, there is scope for this model to be generalised further to be adapted to any light-harvesting system.

7 Statistical and machine learning approaches for prediction of long-time excitation energy transfer dynamics

This chapter is based on Ref [159]: Kimara Naicker, Ilya Sinayskiy, and Francesco Petruccione (arXiv:2210.14160v2).

One of the approaches used to solve for the dynamics of open quantum systems is the hierarchical equations of motion (HEOM). Although it is numerically exact, this method requires immense computational resources to solve. The objective here is to demonstrate whether models such as SARIMA, CatBoost, Prophet, convolutional and recurrent neural networks are able to bypass this requirement. We are able to show this successfully by first solving the HEOM to generate a data set of time series that depict the dissipative dynamics of excitation energy transfer in photosynthetic systems then, we use this data to test the model's ability to predict the long-time dynamics when only the initial short-time dynamics is given. Our results suggest that the SARIMA model can serve as a computationally inexpensive yet accurate way to predict long-time dynamics.

7.1 Introduction

Time series analysis involves methods of analysing a series of data points that are indexed in time order. The objective of the analysis is to collect and study the past observations of a time series to develop an appropriate model which describes the inherent structure of the series. This model is then used to generate future values for the series, i.e. to make forecasts [186]. In this work, the data being analysed is relevant to the dissipative dynamics of excitation energy transfer (EET) in systems similar to the photosynthetic open quantum system regime.

In some cases, information about the underlying dynamical correlations in open quantum systems can be encoded at the initial stages of their evolution. Therefore, it may be possible to obtain long-time dynamics of open quantum systems from the knowledge of their short-time evolution. This conjecture allows the bypass of the need for direct long-time simulations. The simulation of numerically exact methods to describe the dynamics of open quantum systems often require immense computational resources that scale exponentially with the size of the system under study, hence, it is desirable to develop an approach that can accurately predict long-time dynamics of open quantum systems along with eliminating the need for direct calculations to some extent.

Various numerical solutions for the dynamics of open quantum systems have been developed

considering the complexity of system-bath interactions. The dynamics of an open quantum system that are dependent on the Hamiltonian of the system can be described through density matrix-based approaches in the Liouville space of the system. The numerically exact formalism adopted in this study is the hierarchical equations of motion (HEOM) developed by Tanimura and Kubo, and later adapted to biological light harvesting complexes by Ishizaki and Fleming [4, 5, 112]. Machine learning (ML) has been applied to this focus area in many relevant cases [62, 64, 159, 187–189]. L. E. Herrera *et al.* conducted a comparative study where they benchmarked ML models based on their efficiency in predicting long-time dynamics of a two-level quantum system linearly coupled to harmonic bath [189].

Successful time series forecasting depends on an appropriate model fitting. The development of efficient models to improve forecasting accuracy has evolved in literature. A comparison of the predictive capabilities of a standard statistical, an additive regression and a tree-based model against more structurally complex neural network models to simulate the open quantum system dynamics is carried out using Python. The first stage of the dynamics is obtained by solving the HEOM for a sufficiently large theoretical system, thereafter, we train and test suitable models to determine the validity of our approach. That is to predict a time series from that series past values efficiently.

This paper contains several sections which are organized as follows: in Section 7.2 we describe the formalism used, Section 7.3 describes the data pre-processing procedure, Section 7.4 covers the various time series models used, Section 7.5 presents our experimental forecasting results in terms of MSE obtained on relevant datasets and a brief conclusion of our work as well as the prospective future aim in this field.

7.2 The theory and the data

A time series is a sequential set of data points measured over successive times. It is mathematically defined as a set of vectors $x(t), t = 0, 1, 2, \ldots$ where t represents the time elapsed [190]. The variable $x(t)$ can be treated as a random variable. The measurements taken during an event in a time series are arranged in chronological order.

Quantum systems faced in the real world are rarely entirely isolated, hence, it is important to consider the influence of the surrounding environment (bath) when studying the dynamical behaviour of a system. In the process of an open quantum system, such as a photosynthetic pigment-protein complex, evolving over time we can generate a set of time dependent observables that depict the coherent movement of electronic excitations through the system by solving the HEOM. The theoretical background used to generate the data sets used in the study has been discussed in

Section 4.

7.3 Data pre-processing

Building a representative data set is an important first step in every machine learning project. Though the approach developed in this work can be generalized, we discuss the simplest electronic energy transfer system - a dimer (a spin-boson-type model [68]) or two-level system as well as three- and four-level systems where linear chain configurations were imposed. A sample from the data set is shown in Figure 8.
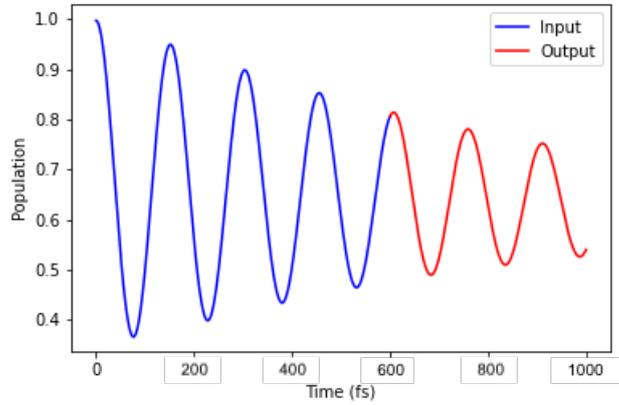
Fig. 8: An example of a sequence generated by the HEOM for a dimer to be split to form the input and output data for the models.

The total system under study can be fully determined by the five independent energy scales: the site energy ϵ_j, the coupling strength J_{jk}, the reorganization energy λ, the cut-off frequency ω_c and thermal energy $k_B T$ of the bath. In order to create a data set suitable for a framework aimed at predicting quantum dynamics in all physically realizable non-Markovian regimes, three parameters were extensively sampled while fixing two parameters: the cut-off frequency $\omega_c = 53$ cm^{-1} and thermal energy $k_B T$ of the bath where $T = 300$ K. These are the typical parameters of photosynthetic EET [45, 191] as seen in Table 5.

Parameter	Lower limit cm^{-1}	Upper limit cm^{-1}
ϵ_j	-100	100
J_{jk}	-100	100
λ	1	100

Table 5: The data set containing time-evolved reduced density matrices is generated for all combinations of the following parameters: the site energy ϵ_j, the coupling strength J_{jk} and the reorganization energy λ.

The HEOM method implemented in Python script is used to solve equation (25). The hierarchy truncation is set to 20 which is a sufficient depth based on the chosen cut-off frequency [142]. To make sense of time series data, it has to be collected over time in the same intervals. The total propagation time is set to 1.0 ps as is sufficient for observing coherent dynamics in photosynthetic EET [142]. For each of the 40 000 samples, observables based on the diagonal elements of the time dependent density matrices i.e. the time evolution of the site populations are collected. Both of the generated observables are divided into shorter trajectories or sequences to test the capability of the models for varying output times to be predicted.

One of our objectives is to determine the trade-off between how far ahead the model can forecast and the shortest input time required to maintain high accuracy in the forecast. As the total propagation time is fixed to 1.0 ps, the actual size of the dataset used varies depending on the lengths of the input and output times. We have performed a grid search in our model testing across varying lengths of input times up to 0.2 ps and output times between 0.01 ps and 0.6 ps.

Before testing, each sample time series was split into multiple shorter slices. For example, a single series split to produce inputs that are 0.2 ps long and outputs 0.6 ps long would generate 1 001 shorter sequences. This sliding window technique is pictured in Figure 9.
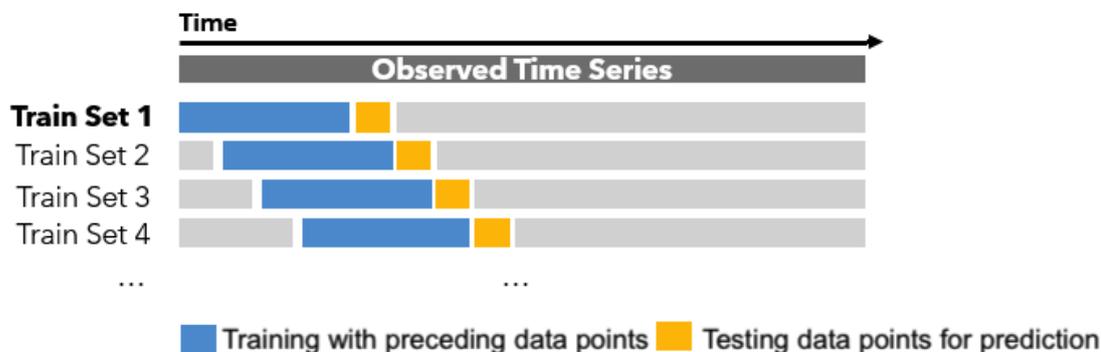
Fig. 9: The use of preceding time steps to predict the following time steps is referred to as the sliding window process. In the figure, the blue portion represents the points in the series that will be used for training and the yellow portion will be used in testing the models. Each training/ testing set differs as the "window" is moved forward over the entire time series by keeping the length of the portions fixed.

The data set is partitioned into a training set of 70% of the data and a validation set of 10% of the data. Additionally, 20% of the data is held out during the training procedure and is used for testing.

7.4 Methodology

In practice, a suitable model is fitted to a given time series and the corresponding parameters of the underlying variables are estimated using the known data values. The procedure of fitting a time series to a proper model is known as Time Series Analysis. It comprises methods that attempt to understand the nature of the series and is often useful for future forecasting and simulation. Past observations are collected and analyzed to build a suitable mathematical model which captures the underlying data generating process for the series. Then in time series forecasting, the future events are predicted using the model [192, 193].

Competent time series analysis remains dominated by traditional statistical methods as well as simpler machine learning techniques such as ensembles of trees and linear fits.

Machine learning is a sub-field of artificial intelligence. Deep learning is a sub-field of machine learning, and neural networks make up the backbone of deep learning algorithms [194]. Machine learning can be broadly defined as computational methods using experience to improve performance or to make accurate predictions. Here, experience refers to the past information available to the learner. Machine learning algorithms can be broadly categorized as unsupervised or supervised by what kind of experience they are allowed to have during the learning process. Supervised learning algorithms experience a data set containing features but each example is also associated with a label or target. In supervised learning, the goal is to learn the relationship between a set of inputs and outputs. Machine learning is intrinsically related to data analysis and statistics considering that the success of a learning algorithm largely depends on the data used. These techniques combine fundamental concepts in computer science with ideas from statistics, probability and optimization. Machine learning in time series analysis dates back many decades. A seminal paper from 1969, "The Combination of Forecasts," analyzed the idea of combining forecasts rather than choosing a best one as a way to improve forecast performance.

Inherently, ensemble methods have set the standard in many forecasting problems. Ensembling rejects the concept of a perfect or even significantly superior forecasting model relative to all possible models.

We aim to estimate which of the models discussed in the following subsections can accurately predict the long-time dynamics of an open quantum system provided the preceding short-time dynamics of the system is known. We have employed the ETNA Time Series Library [195] for constructing and testing the classic SARIMA framework, the CatBoost model and Facebook's in-house model Prophet (specifically designed for learning from business time series) discussed in Sections 7.4.1, 7.4.2 and 7.4.3, respectively. ETNA is a user-friendly time series forecasting framework which includes

built-in toolkits for time series pre-processing and feature generation with a variety of predictive models. By using this package, the process of determining the parameters for the models is automated and optimized. In Section 7.4.4 we discuss neural network approaches for tackling sequential data problems.

7.4.1 SARIMA

We consider conventional statistical methods developed specifically for time series data. The Auto-Regressive Integrated Moving Average (ARIMA) model and its variants account for the correlations that arise between data points in the same time series [196]. This contrasts with the standard methods applied to cross-sectional data in which it is assumed that each data point is independent of the others in the sample. The Auto-Regressive (AR) model relies on the insight that the past predicts the future and so conjectures a time series process in which the value at a point in time t is a function of the series' values at earlier points in time. A Moving Average (MA) model relies on a picture of a process in which the value at each point in time is a function of the past value error terms - each of which is independent from the others. Most time series models require the data to be stationary. This requirement is satisfied when the statistical properties such as mean, variance and covariance remain constant over time. Differencing is a means of removing trends and rendering a time series stationary. It is the process of converting a time series of values into a time series of changes in values over time. The model which combines these along with the added option of differencing is the ARIMA model which recognizes that a time series can have both underlying AR and MA model dynamics.

The series obtained after differencing is the change between consecutive observations in the original series, it can be written as

$$y_t^{'} = y_t - y_{t-1}, \tag{35}$$

where y_t is the tth term in the sequence of T terms. Note that the series in Eq. (35) will have only $T - 1$ values.

The ARIMA model is specified in terms of the parameters (p, d, q). The parameters refer to the auto-regressive, integrated and moving average parts of the data set, respectively. The p parameter is an integer that confirms how many lagged series are going to be used to forecast periods ahead. The d parameter indicates how many differencing orders are going to be used to make the series stationary, if necessary. The q parameter represents the number of lagged forecast error terms in

the prediction equation. SARIMA is a seasonal ARIMA model and it is used for time series with seasonality. The uppercase notation $(P, D, Q)_m$ represent the seasonal components of the model where m is the number of observations per year.

We implemented the SARIMA framework from the publicly available ETNA package. It employs a grid search strategy that determines the optimal parameters for p, d, and q. The predictions on the test set are then obtained by using built-in functions and the results are presented in Section 7.5.

7.4.2 CatBoost

CatBoost, short for category boosting, is an algorithm that is based on decision trees and gradient boosting. Decision trees are supervised learning method used for classification and regression [197]. Gradient boosting is a process of constructing an ensemble predictor by performing gradient descent in a functional space [198]. Boosting is a way of creating an ensemble of models. It begins by fitting an initial model to the data, thereafter another model is constructed that focuses on accurately predicting the cases where the initial model performed poorly. As this process of boosting is repeated, the ensemble is formed which is expected to be better than either model individually. Gradient boosting is taken a step further than this as the next model is combined with previous models to minimise the overall prediction error. The method involves setting the target outputs for the next model based on the gradient of the error with respect to the previous prediction, hence, each subsequent model takes a step in the direction that minimizes prediction error. Although the primary function of this model is for handling categorical data, we find it worth testing whether this model presents any advantages in forecasting. We tested the CatBoost model from the publicly available ETNA package. The results based on the predictions made are presented in Section 7.5.

7.4.3 Facebook Prophet

Facebook's Prophet is an additive regression model designed for making forecasts for uni-variate time series datasets and to automatically find an optimal set of hyperparameters for the model with trends and seasonal structure by default. It belongs to the family of General Additive models (GAM) which fit a set of smooth functions that describe trend, seasonality and predictable special events or cycles to the data. The GAM is the sum of its smooth functions. A GAM treats a time series as a curve-fitting exercise.

Its core comprises the sum of three functions of time plus an error term: growth or trend $g(t)$, seasonality $s(t)$, holidays $h(t)$ and error e_t. The growth function incorporates "changepoints", which are moments in the data where the data shifts direction, to model the overall trend of the data. The

seasonality function is a Fourier Series as a function of time. Prophet automatically detects the Fourier order which is the optimal number of terms in the series. The holiday function allows Prophet to adjust accordingly to an anomaly that may change the forecast [199]. The error term stands for random fluctuations that cannot be explained by the model.

We used the publicly available ETNA implementation of Prophet. The data had to be transformed for use of the model as the input data must contain two fields. Each data point in the time series had to be allocated a date (this must be a valid calendar date from which the holidays can be computed). We allocated a consecutive set of dates to each data point. The second field is the target variable which represents the value to be predicted. The results obtained by using built-in fitting and forecasting functions are presented in Section 7.5.

7.4.4 Neural networks

Artificial neural network (ANN) is the broadest term used to classify a machine that mimics human intelligence. Similar to the human brain, ANNs attempt to recognize regularities and patterns in data, learn from experience and provide generalized results. They are data driven and self-adaptive in nature. It is not necessary to make any a priori assumption about the statistical distribution of the data; the desired model is adaptively formed based on the features presented from the data. This approach is useful for many practical situations, where no theoretical guidance may be available. Although the development of ANNs was mainly biologically motivated, they have been applied in various focus areas for forecasting and classification purposes.

Many of the steps of pre-processing data to fit a model's assumptions are bypassed when neural networks are used - there is no requirement of stationarity, there is no need to develop the art and skill of picking parameters, such as assessing seasonality and order of a seasonal ARIMA model and there is no need to develop a hypothesis about the underlying dynamics of a system, as is helpful with state space modelling. However, this category of algorithms does impose its own pre-processing requirements which we will discuss later.

The first concepts in deep learning brought about three main choices of architecture depending on the data type being handled: Artificial Neural Networks (ANNs) mainly for classification and regression problems; Convolutional Neural Networks (CNN)-based for spatial data (such as images data) and Recurrent Neural Networks (RNN)-based for sequential data (such as time series data). CNNs and RNNs incorporate feature engineering into their framework and eliminate any need to do so manually. This is seen in their capacity to extract features and create informative representations of time series automatically.

The neural network models discussed in Sections 7.4.5 and 7.4.6 have been built, trained and tested using the Scikit-learn [200] and TensorFlow [201] Python packages.

7.4.5 LSTM

At a basic level, a neural network — of which recurrent neural networks (RNNs) are one type, among other types such as convolutional networks (discussed in Section 7.4.6) — is composed of three primary components: the input layer, the hidden layers and the output layer. Each layer consists of nodes or neurons. Feed-forward neural networks (FFNNs), the original single-layer perceptron, developed in 1958 is the precursor to recurrent neural networks. In FFNNs, the information flows in only one direction: from the input layer, through the hidden layers, to the output layer but never backwards in feedback loops.

A recurrent neural network, by contrast, retains a memory of what it has processed. Recurrent neural networks are a broad class of networks specifically designed for processing sequential data. They are attractive to us in this study because of their characteristic capabilities of analysing temporal data as they can retain their state from one iteration to the next by using their own output as input for the next step. In other words, RNNs are a type of neural network that makes recurrent connections by going through temporal feedback loops, thus they can learn from previous iterations during its training. The loops make it a recurrent network.

The hidden layers are placed between the input and the output layer. In an RNN, an output is produced but is also fed back through *backpropagation* as an input for training the hidden layer on the next observation. RNNs carry out the training process by adjusting the weights throughout the neural network. The network recalibrates the weights for both the current and the previous inputs, multiplies the vector of input values with the vector of new weights - this step either enhances or diminishes the importance of each input with respect to the training goal of lowering the prediction error - and lastly, passes the vector of results on as an input to the next layer. By adapting the weights, the hidden layer incrementally derives a function which transforms the input values to output values that approximate the actual observations in the training dataset. However, the function that maps the inputs to the outputs is not expressed as a closed-form equation i.e. it remains hidden.

We will discuss the main building blocks of RNNs. All RNNs have the form of a chain of repeating modules of neural network layers. In standard RNNs, this repeating module will have a simple structure. Within hidden layers, the receiving node calculates a weighted sum of the inputs it receives (which are the output of the preceding layer and the new input) to calculate the total input for the activation (transfer) function. The activation function determines how much the node will contribute

57

to the next layer which is what the output represents. Among the types of frequently used activation functions are the sigmoid function, hyperbolic tangent function (tanh), the step function and the ReLU (rectified linear unit) function.

When the network generates prediction values after a forward pass, the prediction error is also computed which quantifies the deviation from the training dataset through a cost function (also referred to as a loss, error or objective function). The network aims to minimize the error by adjusting its internal weights during training. Backpropagation calculates the partial derivatives of the error with respect to the weights. In each iteration, the RNN recalibrates the weights, up or down, based on the partial derivatives.

The concept of gradient descent refers to the search for a global minimum by evaluating the partial derivatives. The partial derivative with respect to a certain weight reveals how that particular weight contributes towards the total error. The network varies a single weight and records its effect on the total error to obtain its gradient. The repeated adjustment of the weights, down the descent towards a minimal error, will move the model towards an incrementally reduced prediction error. This is computationally intensive hence, the long training phases a RNN typically requires. Gradient descent denotes the search for the global minimum, the set of weights that will minimize the total error.

The RNN updates the old weights by subtracting from them a fraction of their respective gradients. The fraction represents the learning rate, a value above 0 and up to 1 i.e. $newweight = oldweight - gradient * learningrate$. A higher learning rate can speed up the training process of the RNN, although it can also cause overshooting which may prevent the network from settling on a minimal total error. The goal of the RNN is to minimize the cost function. Prediction accuracy metrics such as the mean squared error MSE or root mean squared error RMSE can serve as cost functions.

An epoch represents the feeding of the entire training dataset through the network, consisting of one forward and one backward pass. The number of epochs will determine the trade-off between the time required to train the RNN and its accuracy.

A simple recurrent neural network works well only for a short-term memory. In practice, RNNs suffer from a fundamental problem where they are sometimes unable to capture longer time dependencies in the data. The problem was explored in depth by Hochreiter [202] and Bengio *et al.* [203], who found fundamental justifications for why this caveat exists.

Long Short-Term Memory networks (LSTMs) are a variation of RNNs explicitly designed to avoid the long-term dependency problem. They were introduced by Hochreiter and Schmidhuber [204]. LSTMs

process elements one at a time with memory thus, they are perceived to be more suitable for extracting long range temporal dependencies. Prior to LSTMs, RNNs were "forgetful". Meaning that they could retain a memory although, only pertaining to the process steps in their immediate past. The LSTM, by contrast, introduces loops that can generate long-duration gradients. It can hold on to long-term patterns it discovers while going through its loops. At each time step, it can collate three pieces of information: the current input data, the short-term memory it receives from the preceding cell i.e. the hidden state and the long-term memory from more remote cells (the so-called cell state), from which the RNN cell produces a new hidden state. LSTMs also have a chain like structure, however, the repeating module has a different and more complex structure as seen in Figure 10.

The long-duration gradients resolved a problem termed vanishing gradient descent which occurs when the model stops learning because the gradient's slope becomes too shallow for the search to further improve the weights. This can occur when many of the values involved in repeated gradient calculations are smaller than 1. On the contrary, exploding gradients arises when many values exceed 1 in the repeated matrix multiplications the RNN carries out. The vanishing gradient problem limits an RNN's memory to short-term dependencies whereas the LSTM's architecture keeps the gradients steep enough so that the search does not get "stuck".

A cell in the LSTM is said to be "gated". The cell selectively adds or removes information through the gates as it determines how much incoming information is captured and how much of it is retained. The model can decide whether it opens an input gate to store information, reject and delete it from long-term memory (forget gate) or passes the information on to the next layer (output gate). The RNN is able to make these decisions based on the weights it learns to assign in the process of minimising the error. The gates carry out matrix multiplications between the information values they receive as their current inputs, from short-term or long-term memory. Over time, the LSTM learns which information pieces are effective in reducing the prediction error. It will open and close gates accordingly by assigning higher or lower weights between 0 and 1 to the information values. Through its loops, it will let the useful values, with higher weights, pass through the output gate to form a new short-term memory and it will discard the low-weighted values.

The number of neurons and layers are user definable. Although decreasing the number of neurons and hidden layers can improve the computation time per iteration, this comes at the cost of accuracy. The model also has an added caveat of either underfitting or overfitting. In the case of overfittng, the model will perform well on training data whereas this performance will deteriorate on unseen test data. In order to develop a reliable mode, the model architecture needs to be determined by optimization possibly by a trial-and-error procedure. Many recent works have exhibited the capability of neural
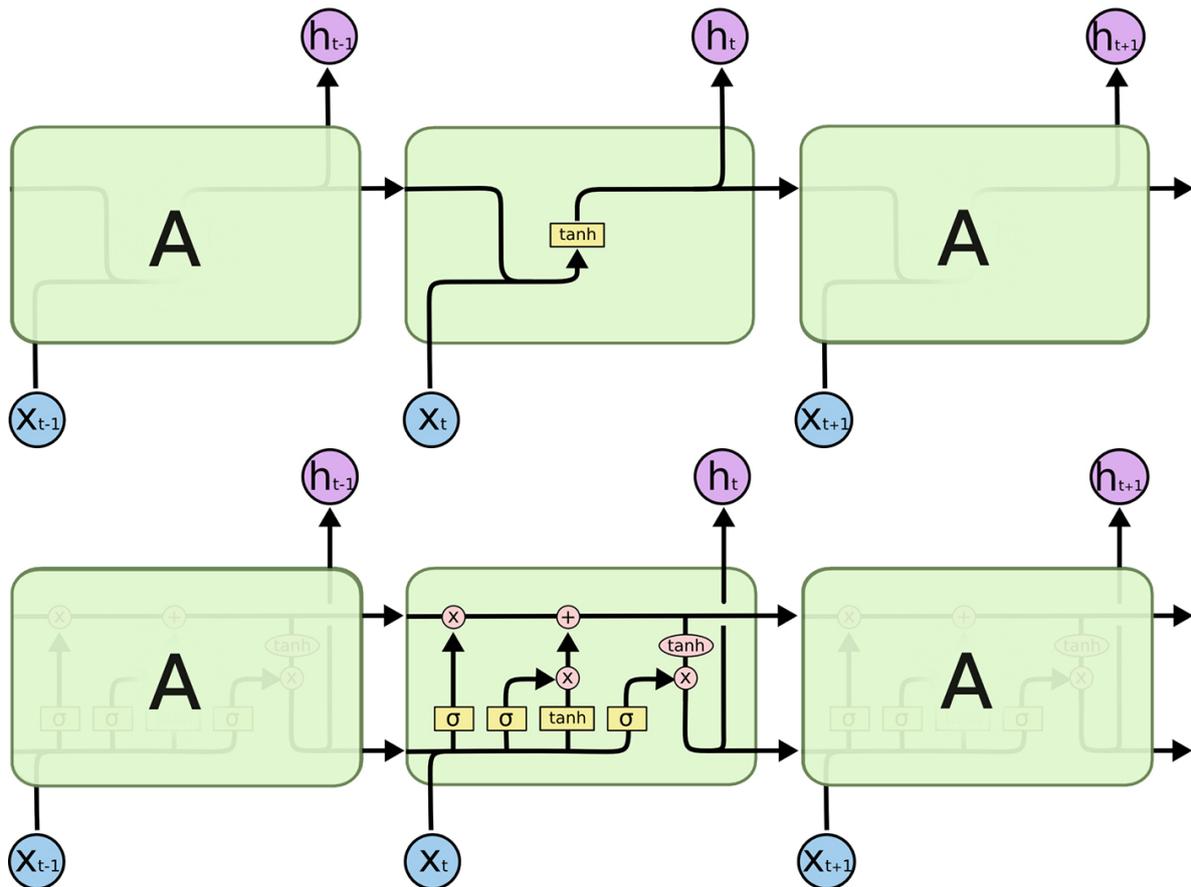
Fig. 10: RNN cell structure (top) vs. LSTM cell structure (bottom). Each line carries an entire vector, from the output of one node to the inputs of others. The pink circles represent pointwise operations, like vector addition, while the yellow boxes are learned neural network layers. Lines merging denote concatenation, while a line forking denote its content being copied and the copies going to different locations [205].

networks, specifically recurrent neural networks (RNNs), for sequential data analysis [206–210].

Other RNN variants — and even other types of LSTMs — exist such as gated recurrent unit networks, bidirectional RNNs and convolutional RNNs which may be used in further testing and implementation. It is also important to note that studies find that the RNN variants do not consistently outperform one another, rather performance is reliant on the data to some extent [211]. There does not seem to be the best RNN variant.

The training and testing procedure for LSTMs is straightforward through the Scikit-learn package. Once the data is loaded in, pre-processed and split into training and validation datasets, it can be fit with the model. The main set of hyperparameters that are needed to set up the model before fitting are the number of input nodes, hidden layers and their nodes, output nodes, number of epochs and the learning rate. Their optimal settings are not known a priori, rather they are problem specific. The

fitting or training process is time-consuming and dependent on the computer's processor performance. Once the model is fitted after training and validation, it can be tested on the test dataset. The results of this stage of testing of our model is presented in the results of Section 7.5.

7.4.6 CNN

It has been shown that using CNNs for time series classification has several important advantages over other methods. They are highly noise-resistant models and they can extract informative features which are independent of time [212].

Consider a time series of length n and width k. The length is the number of time steps and the width is the number of variables in a multi-variate time series. Generally, the first layer in a CNN network is the convolutional layer. This is the core building block and does most of the computational heavy lifting. The convolution kernels have the same width as the time series and their length can vary. The kernel moves in one direction from the beginning of a time series towards its end, performing the convolution operation. The elements of the kernel get multiplied by the corresponding elements of the time series that they cover at a given point (referred to as the receptive field). The results of the multiplication are added together and a non-linear activation function is applied to the value in the activation layer. The most frequently used non-linearity for CNNs is the rectified non-linear unit (ReLU) function which combats the vanishing gradient problem occurring in the sigmoid function. The resulting value becomes an element of a new "filtered" uni-variate time series. The kernel then moves forward along the time series to compute the next value. The number of new "filtered" time series is equivalent to the number of convolution kernels. Depending on the length of the kernel, different characteristics of the initial time series get captured in each of the new filtered series. A representation of this layer can be seen in Figure 11.

Next is the pooling layer. This involves the downsampling of features by applying the max-pooling to each of the filtered time series vectors, the largest value is taken from each vector based on the specified stride and size of the window. A new vector is formed from these values and this vector of maximums is the final feature vector that can be used as an input to a regular fully connected layer. Lastly is the fully connected layer, this involves flattening. In this layer, the entire pooled matrix is transformed into a single column which is then fed to the artificial neural network layers for processing.

There are several significant differences in the working nature of CNNs versus RNNs that make each or the other better suited for a task. CNNs use convolution operations that can handle spatial information available in images. CNNs are computationally cheaper than RNNs as a CNN learns in batches while RNNs train sequentially, as a consequence, a caveat of using an RNN is that parallelization cannot be

done. Unlike RNNs, CNNs learn patterns within a specified range of steps or time window without
the assumption that history is complete. This property can make CNNs attractive for working with
missing data. Data pre-processed for CNNs can be shuffled which allows the CNN to interpret the
data from a broader perspective to a certain extent. RNN models are limited to learning from data in
preceding time steps only. For the case of training a model such that it is dependent on the history of
the data and/ or such that it can handle varying sizes of input and outputs, RNNs are more suitable.

Similarly to RNNs, the training and testing procedure for CNNs is straightforward through the Scikit-
learn package. Once the data is loaded in, pre-processed and split into training and validation datasets,
it can be fitted with the model. The main set of hyperparameters that are needed to set up the model
before fitting are different for each of the three types of layers. For the convolutional layer, the number
and size of kernels and the activation function. The pooling layer requires the stride and size of the
window and the fully connected layers depend on the number of nodes and activation function. Once
the model is fit after training and validation, it can be tested on the test dataset. In terms of CNNs,
it is worth mentioning that the proposed method is not the only one that exists. There are ways of
presenting time series in the form of images to which a regular 2D convolution can be applied. The
results of this stage of testing of our model is presented in the results of Section 7.5.

7.5 Results and conclusion

In this section we report results obtained with the five different approaches to forecasting – namely
SARIMA, CatBoost, Prophet, CNN and LSTM. All training and testing of models was carried out
using a workstation with the following specifications: 2x Intel Xeon E5-2640 2.0GHz, 128 GB RAM.
Maintaining consistency in the machine used is crucial as we are comparing the time required by each
model as well. It is worth noting the difference in meaning of training for the neural networks versus
the rest of the models. In the case of neural networks, in the training stage the model is trained on *all*
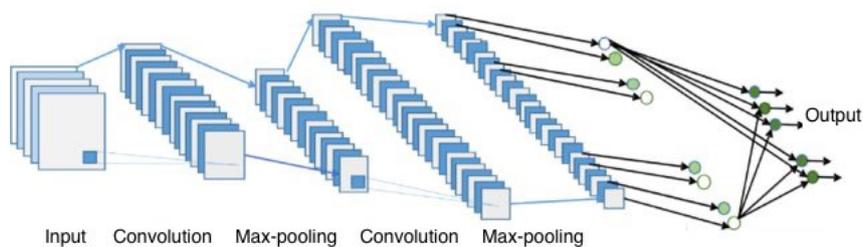

Fig. 11: Vanilla convolutional neural network consisting of convolutional and max-pooling layers fed into a fully-connected
network.

the samples in the training dataset. The weights and biases of the models are adjusted in each epoch until the minimum MSE is achieved. Only after this is the model used for predictions on the test dataset - it is the time taken to train and test the CNN and LSTM models and the MSE achieved on the test dataset that is reported here. On the contrary, SARIMA, CatBoost and Prophet do not perform batch training and testing. The models train on a single sample time series and make predictions based on this single sample training only. Hence, it is the average MSE and time taken *per sample* that is reported in this case.

The accuracy of a models predictions depend on the length of the output time to be predicted as can be expected. Tables 6 and 7 capture the mean squared error results of each model for varying forecasting lengths. We choose an input time of 0.2 ps in all calculations reported in this study as it provides the best compromise between accuracy and computational cost.

Model	2 level		3 level		4 level	
	MSE (10^{-3})	Time (s)	MSE (10^{-3})	Time (s)	MSE (10^{-3})	Time (s)
SARIMA	0.3116 ± 0.0011	0.10	0.3193 ± 0.0056	0.12	0.3491 ± 0.0002	0.12
CatBoost	3.5634 ± 0.0155	4	2.1060 ± 0.0143	5	2.7800 ± 0.0244	7
Prophet	0.1281 ± 0.0099	23	0.1906 ± 0.0100	26	0.1821 ± 0.0140	27
CNN	1.3012 ± 0.0343	835	12.616 ± 0.0254	900	16.010 ± 0.0233	890
LSTM	0.0461 ± 0.0090	7200	0.0923 ± 0.0109	7400	0.1204 ± 0.0100	7500

Table 6: The MSE values and time to train/ test for each model to predict 100 time steps ahead i.e. 0.02 fs.

Model	2 level		3 level		4 level	
	MSE (10^{-3})	Time (s)	MSE (10^{-3})	Time (s)	MSE (10^{-3})	Time (s)
SARIMA	9.7242 ± 0.0121	0.27	12.448 ± 0.0021	0.31	11.306 ± 0.0123	0.37
CatBoost	13.596 ± 0.1098	7	19.558 ± 0.1138	9	14.855 ± 0.0198	8
Prophet	236.66 ± 0.3345	24	307.47 ± 0.4341	22	238.76 ± 0.3971	30
CNN	22.681 ± 0.1143	4200	65.123 ± 0.1909	3990	66.321 ± 0.0967	4300

Table 7: The MSE values and time to train/ test for each model to predict 3 000 time steps ahead i.e. 0.6 fs.

The mean-squared error (MSE) provides a direct quantitative check of the extent to which the predicted response value for a given observation is close to the true response value for that observation. This measure works well in ensuring that our trained model has no outlier predictions with large errors since it allocates larger weight to these errors due to the squaring part of the function. To calculate the MSE as given in Eq. (32), the difference between the models predictions y_i and the ground truth f_i are

squared then, the average across the whole dataset is taken. An ideal MSE value is 0.0, which means that all predicted values matched the expected values exactly. MSE, which squares the prediction errors, penalizes larger errors more than mean absolute percentage error (MAPE) does. Bias arises when the distribution of residuals is left-skewed or right-skewed. The mean will lie above or below the median. A forecast that minimizes the MSE will exhibit less bias.

We emphasize that only the short-time initial dynamical information is required - 1 000 time steps or 0.2 ps. The rest of the time evolution is simulated by reconstructing the observable based on predictions for each time step beyond the initial input time. Thus, if a single-step prediction error is not sufficiently small then the error will rapidly accumulate resulting in a deterioration of the accuracy. As seen in the reported results, all models are able to reproduce long-time dynamics nearly exactly and irrespective of the dynamical regime. The LSTM model was excluded from the training and testing at the longer time due to its exhaustive computational requirements i.e. the trade-off between accuracy and time did not seem valid here.

With regards to hyperparameter settings, we have limited the learning rate, 0.001 and the epochs, 300, in our setup of the LSTM and CNN models. A tuning algorithm could tweak them while running the fitting process to try to achieve an even lower MSE. On the other hand, this particular dataset of time series is not greatly complex as it is purely dissipative. It does not seem likely that time-intensive tuning efforts would reduce the MSE much further.

The ability of each model to learn the fundamental properties of the density matrix is tested on the unit trace and positive semi-definite properties. We emphasize that these properties are not enforced during the training. We found that all models were able to learn both properties, however, the SARIMA model proved to be the fastest, most efficient and most accurate. An observation to note is that each model produced roughly the same results for each level of systems. It is also evident how the complexity of the HEOM increases with increasing system size i.e. more computational power and time required to explicitly solve the set of equations. The importance of our results is that these machine learning models provide a cheaper yet still greatly efficient alternative to determining the long-time dynamics of these systems. This could be widely used depending on the level of accuracy required for a particular use case.

In Figure 12, we present the results from testing the predictive capability of the SARIMA model on the Fenna–Matthews–Olson (FMO) protein of green sulfur bacteria [20]. The time evolution of the population of the first three sites of the complex are used for training and testing where site one is the initial excited state for numerical calculations. The FMO complex was one of the first systems of its kind in which long-lived quantum coherence was observed which motivated the development of a

64

new quantum dynamic equation for excitation energy transfer i.e. HEOM, hence, the relevance and importance of testing our model on this data set.
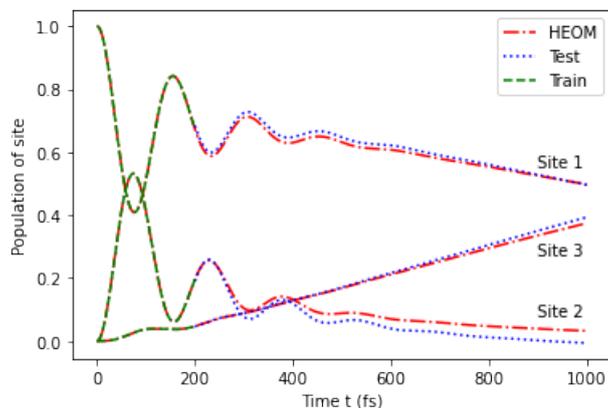

Fig. 12: Comparison of HEOM results (red) with predictions made by SARIMA (blue) based on short-time training data (green) for sites 1-3 of the well-known photosynthetic seven-site pigment protein Fenna–Matthews–Olson complex.

Furthermore and as anticipated during testing, the models performed slightly worse in the region of small reorganization energy and low temperature which is the regime of weakly damped coherent oscillatory dynamics. This is largely due to the input data being almost indistinguishable in this regime. In summary, we demonstrate that SARIMA, CatBoost, Prophet, convolutional and recurrent neural networks models can predict the long-time dynamics of an open quantum system provided the preceding short-time dynamics of the system is known. However, owing to its efficiency with regards to computational power required and time required to conduct simulations plus, low MSEs achieved, the SARIMA appears to be the model that satisfies our criteria. The model has been trained on a data sets relevant to photosynthetic excitation energy transfer and we aim to scale and deploy it to investigate long-lasting quantum coherent phenomena observed in larger light-harvesting complexes. The approach is a practical tool in terms of reducing the required computational resources for long-time simulations whilst maintaining high accuracy. A photosynthetic EET was chosen as an example, however, the same approach can be used to study other phenomena provided the relevant data sets are available.

8   Discussion and conclusion of thesis

This thesis investigated how classical machine learning can be leveraged for investigations of a certain class of problems pertaining to dynamics in the regime of excitation energy transfer in open quantum systems. As a conclusion, the central points made in each chapter are discussed, a summary of the main findings is given as well as suggestions on avenues for further research.

Quantum systems encountered in the real world are seldom completely isolated as it is the interaction between an open quantum system and its environment that causes energy dissipation and usually leads to the destruction of coherence [77]. Machine learning techniques can be applied to investigate the theory of open quantum systems in further detail. This work focuses more specifically on leveraging classical machine learning models to study the dynamics of excitation energy transfer (EET) within these complex open systems.

There exist numerous studies of machine learning techniques applied to accelerate computations by many orders of magnitude at a reasonable level of accuracy in a bid to understand the temporal evolution of open quantum systems further [21, 22, 65, 132, 177, 178]. Insight into long-lived quantum coherence in EET processes can be gained through the reduced system dynamics which are described by the reduced density operator [77]. There exist conventional theories that have been developed to simulate reduced dynamics including the numerically exact hierarchical equations of motion (HEOM) method which is used in this work [6, 74, 138–140].

The motivation for this study follows from the first step of photosynthesis in light harvesting complexes where energy is absorbed by the antenna pigments and subsequently transferred to the reaction center in which photochemical reactions are initiated. This process can be modeled by EET from an initially excited pigment to an acceptor pigment. One of the possible explanations for the EET in light harvesting systems is the quantum mechanical nature of the transfer process. The energy of the absorbed sunlight simultaneously travels along all possible paths to the reaction centre by exploiting quantum superposition. Typical situations in photosynthetic EET are such that the electronic coupling strengths, between chromophores and their local environment phonons, span a similar range as the reorganization energies, which characterize the time scale of the coupled phonons relaxing to their respective equilibrium states [132, 133].   These site-dependent reorganization processes can be described by theories derived in the non-Markovian regime. Therefore, we employ the HEOM framework which can describe quantum coherent wavelike motion and incoherent hopping in the same framework and reduces to the conventional Redfield [74, 138, 139] and Förster [140, 141] theories in their respective limits of validity.

One may solve the HEOM to study the excited state population dynamics of a given complex. Although these calculations are highly computationally expensive in terms of time, storage, and computational power due to the adverse scaling with the number of sites in a given complex, they can still be done. In the process of an open quantum system evolving, we are able to generate data that depicts its temporal behaviour. This data is usually in the form of sequences of two-dimensional time dependent spectroscopic data and is used for two different purposes in this work.

Two main problems were addressed in this thesis. The first being the inverse problem of the HEOM and the second focusing on time-series analysis/prediction. The outcomes of both the research problems investigated are positive and show promising results to further motivate future work in these areas. We will discuss the two research problems separately here.

The first objective is to determine whether a trained machine learning model can accurately describe the complex under study when given this time dependent data, i.e. the inverse problem. Machine learning of temporal and spatial data has been an active area of research in physics due to the wide span of potential applications [213–215]. The task here has been to develop a scalable and efficient tool for the description of the dynamical and thermodynamical properties of open quantum systems by use of a trained convolutional neural network (CNN) as CNNs allow for extraction of features in 3 dimensions. The trained CNN has the task of providing the solution to this mathematically challenging inverse problem. In this approach, the model is trained and tested on a large and diverse enough dataset of Hamiltonians and corresponding time evolution of reduced density matrices such that it may learn patterns in the data and be able to present highly accurate predictions of system Hamiltonian parameters without having any knowledge of the theory or in this case, HEOM. We explicitly solve the HEOM to generate training and testing datasets for supervised machine learning tasks where elements of reduced density matrices are translated into features for the model and corresponding excited state energies and electronic couplings are used as labels.

The parameters of the Hamiltonians in these datasets are motivated by and sampled around the same order of magnitude as those that are typical of light-harvesting pigment-protein complexes. The datasets generated for training of CNNs are sampled from a range of 12 200 to 12 600 cm^{-1} for excited state energies and -100 to 100 cm^{-1} for electronic couplings as it was found that it is within these ranges that the coherence lasts for longer timescales. The datasets were generated to simulate excitonic systems with more than two sites or pigments only to demonstrate the adaptability and applicability of the CNN to make accurate predictions for more multifaceted systems. The developed models have been able to make predictions with up to 99.28% accuracy.

This work can be developed further in a few ways which we will now discuss. One possible route

would be to investigate higher dimension convolutional neural networks. These models may have the ability to achieve higher accuracy predictions and more importantly, they might be able to overcome the caveat we came across in the overdamped regime. This is the regime where the time dependent observables display a purely dissipative behaviour in which our developed model performed poorly. An alternate class of models worth investigating is that of physics-informed neural networks (PINNs): A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations proposed by Raissi *et al.* [216]. Jagtap *et al.* recently solved inverse problems in supersonic flows using PINNs, and extended PINNs (XPINNs) [217] methods [218]. Furthermore, we intend on developing the next model in terms of transfer learning such that it may be independent of the dimension of the input data. In the current set-up, for a two-level system, two observables are input and for a three-level system, three observables are input and so on. This modification would allow the user to input observable data such that in return the model may determine the dimension of the system under study which represents the number of pigments in the complex. This would mean that it may not be necessary for the structure of the system under study i.e. the number of sites, to be known. With regards to the data we input to the model, we may also test the effects of incorporating some information regarding the environment surrounding the system. This could be in the form of bath-relaxation time, correlation functions or spectral density functions amongst other characteristics. On the other hand, we may modify the model to predict parameters of the spectral density along with performing Hamiltonian tomography.

Secondly, in terms of objectives in this study, the simulation of the dynamics of quantum dissipative systems such as those primarily focused on in this work is a challenging problem across physics and chemistry. There exists a multitude of numerically exact methods that have been developed for quantum dynamics simulations including the HEOM which we have been able to solve and generate datasets based on these to a 'shallow' depth of the hierarchy. However, the preeminent demand in this approach is the computational power required in order to solve the HEOM to much greater depths. Machine learning offers alternative methods to accurate, yet greatly accelerated quantum dynamics calculations with minimum input information required.

We demonstrate that a machine learning model can predict the long-time dynamics of an open quantum system provided the preceding short-time dynamics of the system is known. The model has been trained on a datasets relevant to photosynthetic excitation energy transfer and we aim to scale and deploy it to investigate long-lasting quantum coherent phenomena observed in larger light-harvesting complexes. The approach is a practical tool in terms of reducing the required computational resources for long-time simulations whilst maintaining high accuracy. Our results

from testing the predictive capability of the SARIMA model on the Fenna–Matthews–Olson (FMO) protein of green sulfur bacteria shown in Figure 12 provides a visual comparison of the high accuracy the model is capable of. The FMO complex was one of the first systems of its kind in which long-lived quantum coherence was observed which motivated the development of new a quantum dynamic equation for excitation energy transfer i.e. HEOM, hence, the relevance and importance of testing our model on this data set.

Many popular machine learning methods have been studied in this chapter. The field of artificial intelligence and machine learning is growing at a fast pace, hence, future work will focus on novel approaches to time-series modeling.

With regards to the data set we generated to carry out our investigations, it can be modified or broadened in multiple ways. In the cuurent model, we make use of the time dependence of the diagonal elements of the reduced density matrix as our inputs which represents the population of the energy in each site. We may investigate the effects of using fewer inputs and/or more inputs such as the off-diagonal elements which correspond to the coherence in the system. We have limited our electronic coupling energies and reorganisation strengths based on typical ranges of these values estimated for light harvestings complexes such as FMO. We may go beyond these limitations to investigate the dynamics of EET theoretically. The regime we have developed our data in is characterised by strong coupling and non-Markovianity. We can consider the case with weak intersite coupling. In a region of small reorganization energy, the HEOM coincides with that of the Redfield equation. This is where the relaxation of the reduced density matrix is much slower than the correlation time of the phonon-induced fluctuation which characterises the precondition which makes the Markov approximation appropriate for description of the quantum dynamics. It may be worth investigating inverse problems in the regime where the Markov approximation is valid. That is where the phonons are able to relax to their equilibrium states instantaneously i.e. the phonons are always in equilibrium even under the electron-phonon interaction. In this case, we will be able to employ the Redfield equation [74, 138]. Still in the weak electronic coupling case, with increasing reorganization energy, the HEOM deviates from Redfield and the HEOM remains reliable due to its consideration of non-Markovianity. On another note, we may investigate the implications of considering spectral density functions other that the used Drude-Lorentz spectral density.

References

[1] G. S. Engel *et al.*, "Evidence for wavelike energy transfer through quantum coherence in photosynthetic systems," *Nature*, vol. 446, p. 782–786, 2007.

[2] G. Panitchayangkoon *et al.*, "Direct evidence of quantum transport in photosynthetic light-harvesting complexes," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, p. 20908–20912, 2011.

[3] G. Panitchayangkoon, D. Hayes *et al.*, "Long-lived quantum coherence in photosynthetic complexes at physiological temperature," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, pp. 12 766–12 770, 2010.

[4] Y. Tanimura and R. Kubo, "Time evolution of a quantum system in contact with a nearly gaussian-markoffian noise bath," *Journal of the Physical Society of Japan*, vol. 58, p. 101–114, 1989.

[5] Y. Tanimura, "Numerically "exact" approach to open quantum dynamics: The hierarchical equations of motion (heom)," *Journal of Chemical Physics*, vol. 153, p. 020901, 2020.

[6] A. Ishizaki and G. R. Fleming, "Theoretical examination of quantum coherence in a photosynthetic system at physiological temperature," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, pp. 17 255–17 260, 2009.

[7] G. Carleo *et al.*, "Machine learning and the physical sciences," *Reviews of Modern Physics*, vol. 91, p. 045002, 2019.

[8] M. Schuld *et al.*, "An introduction to quantum machine learning," *Contemporary Physics*, vol. 56, p. 172, 2015.

[9] V. Dunjko and H. J. Briegel, "Machine learning & artificial intelligence in the quantum domain: a review of recent progress," *Reports on Progress in Physics*, vol. 81, p. 074001, 2018.

[10] P. Mehta *et al.*, "A high-bias, low-variance introduction to machine learning for physicists," *Physics Reports*, vol. 810, p. 1, 2019.

[11] J. Adolphs and T. Renger, "How proteins trigger excitation energy transfer in the FMO complex of green sulfur bacteria," *Biophysical Journal*, vol. 91, p. 2778–2797, 2006.

[12] C. Kreisbeck *et al.*, "High-performance solution of hierarchical equations of motion for studying energy transfer in light-harvesting complexes," *Journal of Chemical Theory and Computation*, vol. 7, p. 2166–2174, 2011.

[13] M. Sarovar *et al.*, "Quantum entanglement in photosynthetic light-harvesting complexes," *Nature Physics*, vol. 6, p. 462–467, 2010.

[14] M. Mohseni *et al.*, "Environment-assisted quantum walks in photosynthetic energy transfer," *Journal of Chemical Physics*, vol. 129, p. 174106, 2008.

[15] M. B. Plenio and S. F. Huelga, "Dephasing-assisted transport: quantum networks and biomolecules," *New Journal of Physics*, vol. 10, p. 113019, 2008.

[16] S. Shim *et al.*, "Atomistic study of the long-lived quantum coherences in the Fenna-Matthews-Olson complex," *Biophysical Journal*, vol. 102, p. 649–660, 2012.

[17] J. Cao and R. J. Silbey, "Optimization of exciton trapping in energy transfer processes," *Journal of Physical Chemistry A*, vol. 113, p. 13825–13838, 2009.

[18] J. Moix *et al.*, "Efficient energy transfer in light-harvesting systems, iii: The influence of the eighth bacteriochlorophyll on the dynamics and efficiency in FMO," *Journal of Physical Chemistry Letters*, vol. 2, p. 3045–305, 2011.

[19] A. Ishizaki and G. R. Fleming, "On the interpretation of quantum coherent beats observed in two-dimensional electronic spectra of photosynthetic light harvesting complexes," *Journal of Physical Chemistry B*, vol. 115, p. 6227–6233, 2011.

[20] R. E. Fenna and B. W. Matthews, "Chlorophyll arrangement in a bacteriochlorophyll protein from chlorobium limicola," *Nature*, vol. 258, p. 573–577, 1975.

[21] F. Häse *et al.*, "Machine learning for quantum dynamics: deep learning of excitation energy transfer properties," *Journal of Chemical Sciences*, vol. 8, pp. 8419–8426, 2017.

[22] F. Häse, S. Valleau *et al.*, "Machine learning exciton dynamics," *Chemical Science*, vol. 7, pp. 5139–5147, 2016.

[23] S. Valleau *et al.*, "On the alternatives for bath correlators and spectral densities from mixed quantum-classical simulations," *Journal of Chemical Physics*, vol. 137, p. 224103, 2012.

[24] J. M. Olson, "The FMO protein," *Photosynthesis Research*, vol. 80, pp. 181–187, 2004.

[25] R. Pearlstein and R. Hemenger, "Bacteriochlorophyll electronic transition moment directions in bacteriochlorophyll a-protein," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 75, p. 4920, 1978.

[26] D. Hayes and G. Engel, "Extracting the excitonic hamiltonian of the Fenna-Matthews-Olson complex using three-dimensional third-order electronic spectroscopy," *Biophysical Journal*, vol. 100, p. 2043, 2011.

[27] R. Ramakrishnan *et al.*, "Electronic spectra from tddft and machine learning in chemical space," *Journal of Chemical Physics*, vol. 143, p. 084111, 2015.

[28] K. Hansen *et al.*, "Machine learning predictions of molecular properties: Accurate many-body potentials and nonlocality in chemical space," *Journal of Physical Chemistry Letters*, vol. 6, p. 2326–2331, 2015.

[29] M. Rupp *et al.*, "Fast and accurate modeling of molecular atomization energies with machine learning," *Physical Review Letters*, vol. 108, p. 058301, 2012.

[30] E. O. Pyzer-Knapp *et al.*, "Learning from the harvard clean energy project: The use of neural networks to accelerate materials discovery," *Advanced Functional Materials*, vol. 25, p. 6495–6502, 2015.

[31] J. Behler and M. Parrinello, "Generalized neural-network representation of high-dimensional potential-energy surfaces," *Physical Review Letters*, vol. 98, p. 146401, 2007.

[32] S. Caffarri *et al.*, "Functional architecture of higher plant photosystem ii supercomplexes," *The EMBO Journal*, vol. 28, p. 3052–3063, 2009.

[33] N. R. Baker, "Chlorophyll fluorescence: a probe of photosynthesis in vivo," *Annual Review of Plant Biology*, vol. 59, p. 89–113, 2008.

[34] C. Kreisbeck and A. Aspuru-Guzik, "Efficiency of energy funneling in the photosystem ii supercomplex of higher plants," *Journal of Chemical Science*, vol. 7, p. 4174–4183, 2016.

[35] K. Amarnath *et al.*, "Multiscale model of light harvesting by photosystem ii in plants," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 113, p. 1156–1161, 2016.

[36] G. D. Scholes *et al.*, "Lessons from nature about solar light harvesting," *Nature Chemistry*, vol. 3, p. 763–774, 2011.

[37] G. D. Scholes, G. R. Fleming *et al.*, "Using coherence to enhance function in chemical and biophysical systems," *Nature*, vol. 543, p. 647–656, 2017.

[38] C. Kreisbeck and T. Kramer, "Long-lived electronic coherence in dissipative exciton dynamics of light-harvesting complexes," *Journal of Physical Chemistry Letters*, vol. 3, p. 2828–2833, 2012.

[39] A. W. Chin *et al.*, "The role of non-equilibrium vibrational structures in electronic coherence and recoherence in pigment–protein complexes," *Nature Physics*, vol. 9, p. 113–118, 2013.

[40] N. Christensson *et al.*, "Origin of long-lived coherences in light-harvesting complexes," *Journal of Physical Chemistry B*, vol. 116, p. 7449–7454, 2012.

[41] J. C. Dean *et al.*, "Vibronic enhancement of algae light harvesting," *Chemistry*, vol. 1, p. 858–872, 2016.

[42] E. Romero *et al.*, "Quantum coherence in photosynthesis for efficient solar energy conversion," *Nature Physics*, vol. 10, p. 676–682, 2014.

[43] A. D. Sio *et al.*, "Tracking the coherent generation of polaron pairs in conjugated polymers," *Nature Communications*, vol. 7, p. 13742, 2016.

[44] E. Collini *et al.*, "Coherently wired light-harvesting in photosynthetic marine algae at ambient temperature," *Nature*, vol. 463, pp. 644–647, 2010.

[45] T. Brixner *et al.*, "Two-dimensional spectroscopy of electronic couplings in photosynthesis," *Nature (London)*, vol. 434, p. 625–628, 2005.

[46] J. Schulze and O. Kühn, "Explicit correlated exciton-vibrational dynamics of the FMO complex," *Journal of Physical Chemistry B*, vol. 119, p. 6211–621, 2015.

[47] B. Hein *et al.*, "Modelling of oscillations in two-dimensional echo-spectra of the Fenna–Matthews–Olson complex," *New Journal of Physics*, vol. 14, p. 023018, 2012.

[48] D. Suess *et al.*, "Hierarchy of stochastic pure states for open quantum system dynamics," *Physical Review Letters*, vol. 113, p. 150403, 2014.

[49] P. Nalbach *et al.*, "Exciton transfer dynamics and quantumness of energy transfer in the Fenna-Matthews-Olson complex," *Physical Review E*, vol. 84, p. 041926, 2014.

[50] T. Scholak *et al.*, "Optimal networks for excitonic energy transport," *Journal of Physics B*, vol. 44, p. 184012, 2011.

[51] S. Mostarda *et al.*, "Structure–dynamics relationship in coherent transport through disordered systems," *Nature Communications*, vol. 4, p. 2296, 2013.

[52] S. Baghbanzadeh and I. Kassal, "Distinguishing the roles of energy funnelling and delocalization in photosynthetic light harvesting," *Physical Chemistry Chemical Physics*, vol. 18, p. 7459–7467, 2016.

[53] S. Baghbanzadeh *et al.*, "Geometry, supertransfer, and optimality in the light harvesting of purple bacteria," *Journal of Physical Chemistry Letters*, vol. 7, p. 3804–3811, 2016.

[54] V. May and O. Kühn, *Charge and energy transfer dynamics in molecular systems*.   John Wiley & Sons, 2008.

[55] H. S. Bhat *et al.*, "Machine learning a molecular hamiltonian for predicting electron dynamics," *International Journal of Dynamics and Control*, vol. 8, p. 1089–1101, 2020.

[56] F. Häse *et al.*, "Designing and understanding light-harvesting devices with machine learning," *Nature Communications*, vol. 11, p. 4587, 2020.

[57] F. Brockherde *et al.*, "Bypassing the kohn-sham equations with machine learning," *Nature Communications*, vol. 8, p. 872, 2017.

[58] Z. Li *et al.*, "Molecular dynamics with on-the-fly machine learning of quantum-mechanical forces," *Physical Review Letters*, vol. 114, p. 096405, 2015.

[59] M. Krämer *et al.*, "Charge and exciton transfer simulations using machine-learned hamiltonians," *Journal of Chemical Theory and Computation*, vol. 16, pp. 4061–4070, 2020.

[60] C. Wang *et al.*, "Machine learning for predicting electron transfer coupling," *The Journal of Physical Chemistry A*, vol. 123, pp. 7792–7802, 2019.

[61] C. Wang, I. Joanito *et al.*, "Artificial neural networks for predicting charge transfer coupling," *The Journal of Chemical Physics*, vol. 153, p. 214113, 2020.

[62] L. E. H. Rodriguez and A. A. Kananenka, "Convolutional neural networks for long-time dissipative quantum dynamics," *The Journal of Physical Chemistry Letters*, vol. 12, p. 2476–2483, 2021.

[63] D. Wu *et al.*, "Forecasting nonadiabatic dynamics using hybrid convolutional neural network/long short-term memory network," *The Journal of Chemical Physics*, vol. 155, p. 224104, 2021.

[64] K. Lin *et al.*, "Simulation of open quantum dynamics with bootstrap-based long short-term memory recurrent neural network," *The Journal of Physical Chemistry Letters*, vol. 12, pp. 10 225–10 234, 2021.

[65] M. Rodríguez and T. Kramer, "Machine learning of two-dimensional spectroscopic data," *The Journal of Chemical Physics*, vol. 520, pp. 52–60, 2019.

[66] M. A. Nielsen and I. Chuang, *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.

[67] E. Schrödinger, "An undulatory theory of the mechanics of atoms and molecules," *Physical Review*, vol. 28, p. 1049, 1926.

[68] A. J. Leggett *et al.*, "Dynamics of the dissipative two-state system," *Reviews of Modern Physics*, vol. 59, p. 1, 1987.

[69] K. Blum, *Density Matrix Theory and Applications*. New York: Plenum Press, 1996.

[70] A. Nitzan, *Chemical Dynamics in Condensed Phases: Relaxation, Transfer, and Reactions in Condensed Molecular Systems*. New York: Oxford University Press, 2006.

[71] U. Weiss, *Quantum Dissipative Systems*. Singapore: World Scientific, 2008.

[72] G. Lindblad, "On the generators of quantum dynamical semigroups," *Communications in Mathematical Physics*, vol. 48, p. 119–130, 1976.

[73] V. Gorini *et al.*, "Completely positive dynamical semigroups of n-level systems," *Journal of Mathematical Physics*, vol. 17, p. 821, 1976.

[74] A. Redfield, "The theory of relaxation processes," *Advances in Magnetic and Optical Resonance*, vol. 1, pp. 1–32, 1965.

[75] O. M. Scully and E. W. Lamb, "Quantum theory of an optical maser. i. general theory," *Physical Review*, vol. 159, pp. 208–226, 1967.

[76] B. Mollow and M. Miller, "The damped driven two-level atom," *Annals of Physics*, vol. 52, pp. 464–478, 1969.

[77] H. P. Breuer and F. Petruccione, *The theory of open quantum systems*. Oxford University Press, 2002.

[78] E. B. Davies, *Quantum Theory of Open Systems*. New York: Academic Press, 1976.

[79] H. Spohn, "Kinetic equations from hamiltonian dynamics: Markovian limits," *Reviews of Modern Physics*, vol. 52, p. 569, 1980.

[80] P. Pechukas, "Is the dynamics of open quantum systems always linear?" *Physical Review Letters*, vol. 73, p. 1060, 1994.

[81] V. Reimer *et al.*, "Five approaches to exact open-system dynamics: Complete positivity, divisibility, and time-dependent observables," *Journal of Chemical Physics*, vol. 151, p. 044101, 2019.

[82] Y. Tanimura, "Stochastic liouville, langevin, fokker–planck, and master equation approaches to quantum dissipative systems," *Journal of the Physical Society of Japan*, vol. 75, p. 082001, 2006.

[83] Y.Tanimura, "Reduced hierarchical equations of motion in real and imaginary time: Correlated initial states and thermodynamic quantities," *Journal of Chemical Physics*, vol. 141, p. 044114, 2014.

[84] N. Lambert *et al.*, "Quantum biology," *Nature Physics*, vol. 9, pp. 10–18, 2013.

[85] H. B. Chen *et al.*, "Using non-markovian measures to evaluate quantum master equations for photosynthesis," *Scientific Reports*, vol. 5, p. 12753, 2015.

[86] A. Ishizaki and G. Fleming, "Quantum coherence in photosynthetic light harvesting," *Annual Review of Condensed Matter Physics*, vol. 3, p. 333–361, 2012.

[87] P. Strasberg *et al.*, "Nonequilibrium thermodynamics in the strong coupling and non-markovian regime based on a reaction coordinate mapping," *New Journal of Physics*, vol. 18, p. 073007, 2016.

[88] D. Newman *et al.*, "Performance of a quantum heat engine at strong reservoir coupling," *Physical Review E*, vol. 95, p. 032139, 2017.

[89] A. A. Anappara *et al.*, "Signatures of the ultrastrong light-matter coupling regime," *Physical Review B*, vol. 79, p. 201303, 2009.

[90] T. Niemczyk *et al.*, "Circuit quantum electrodynamics in the ultrastrong coupling regime," *Nature Physics*, vol. 6, p. 772–776, 2010.

[91] A. F. Kockum *et al.*, "Ultrastrong coupling between light and matter," *Nature Reviews Physics*, vol. 1, pp. 19–40, 2018.

[92] P. Forn-Díaz *et al.*, "Ultrastrong coupling regimes of light-matter interaction," *Reviews of Modern Physics*, vol. 91, p. 025005, 2019.

[93] I. de Vega and D. Alonso, "Dynamics of non-markovian open quantum systems," *Reviews of Modern Physics*, vol. 89, p. 015001, 2017.

[94] L. Magazzù *et al.*, "Probing the strongly driven spin-boson model in a superconducting quantum circuit," *Nature Communications*, vol. 9, p. 1403, 2018.

[95] W. M. Zhang *et al.*, "General non-markovian dynamics of open quantum systems," *Physical Review Letters*, vol. 109, p. 170402, 2012.

[96] M. Yang and G. R. Fleming, "Influence of phonons on exciton transfer dynamics: comparison of the redfield, förster, and modified redfield equations," *Chemical Physics*, vol. 275, p. 355, 2002.

[97] M. Ban *et al.*, "Reduced dynamics and the master equation of open quantum systems," *Physics Letters A*, vol. 374, pp. 2324–2330, 2010.

[98] L. Mazzola *et al.*, "Pseudomodes as an effective description of memory: Non-markovian dynamics of two-state systems in structured reservoirs," *Physical Review A*, vol. 80, p. 012104, 2009.

[99] J. Iles-Smith *et al.*, "Environmental dynamics, correlations, and the emergence of noncanonical equilibrium states in open quantum systems," *Physical Review A*, vol. 90, p. 032114, 2014.

[100] N. Lambert *et al.*, "Modelling the ultra-strongly coupled spin-boson model with unphysical modes," *Nature Communications*, vol. 10, p. 3721, 2019.

[101] J. Bretón *et al.*, "Relaxation of quantum systems weakly coupled to a bath. i. total-time-ordering-cumulant and partial-time-ordering-cumulant non-markovian theories," *Physical Review A*, vol. 30, p. 542, 1984.

[102] S. Mukamel, "Non-markovian theory of molecular relaxation. i. vibrational relaxation and dephasing in condensed phases," *Chemical Physics*, vol. 37, p. 33, 1979.

[103] E. Schrödinger, *What is life?* Cambridge University Press, 1944.

[104] P. Jordan, *Die Physik und das Geheimnis des organischen Lebens.* Friedrich Vieweg & Sohn, 1941.

[105] L. Turin, "A spectroscopic mechanism for primary olfactory reception," *Chemical Senses*, vol. 21, p. 773, 1996.

[106] T. Ritz *et al.*, "Resonance effects indicate a radical-pair mechanism for avian magnetic compass," *Nature*, vol. 429, p. 177, 2004.

[107] R. E. Blankenship *et al.*, "Comparing photosynthetic and photovoltaic efficiencies and recognizing the potential for improvement," *Science*, vol. 332, p. 805, 2011.

[108] F. Caruso *et al.*, "Highly efficient energy excitation transfer in light-harvesting complexes: The fundamental role of noise-assisted transport," *Journal of Chemical Physics*, vol. 131, p. 105106, 2009.

[109] P. Rebentrost *et al.*, "Role of quantum coherence in chromophoric energy transport," *Journal of Physical Chemistry B*, vol. 113, p. 9942–9947, 2009.

[110] M. Mohseni *et al.*, "Geometrical effects on energy transfer in disordered open quantum systems," *Journal of Chemical Physics*, vol. 138, p. 204309, 2013.

[111] A. W. Chin *et al.*, "Exact mapping between system-reservoir quantum models and semi-infinite discrete chains using orthogonal polynomials," *Journal of Mathematical Physics*, vol. 51, p. 092109, 2010.

[112] A. Ishizaki and G. R. Fleming, "Unified treatment of quantum coherent and incoherent hopping dynamics in electronic energy transfer: Reduced hierarchy equation approach," *Journal of Chemical Physics*, vol. 130, p. 234111, 2009.

[113] B. W. Matthews *et al.*, "Structure of a bacteriochlorophyll-a protein from the green photosynthetic bacterium prosthecochloris aestuarii," *Journal of Molecular Biology*, vol. 131, p. 259–285, 1979.

[114] D. E. Tronrud *et al.*, "Structure and x-ray amino acid sequence of a bacteriochlorophyll-a protein from prosthecochloris aestuarii refined at 1.9 resolution," *Journal of Molecular Biology*, vol. 188, p. 443–454, 1986.

[115] Y. F. Li *et al.*, "Crystal structure of the bacteriochlorophyll-a protein from chlorobium tepidum," *Journal of Molecular Biology*, vol. 271, p. 456–471, 1997.

[116] F. Müh *et al.*, "Alpha-helices direct excitation energy flow in the Fenna–Matthews–Olson protein," *Proceedings of the National Academy of Sciences USA*, vol. 104, pp. 16 862–16 867, 2007.

[117] H. van Amerongen *et al.*, *Photosynthetic Excitons*. Singapore: World Scientific, 2000.

[118] R. Blankenship, *Molecular Mechanisms of Photosynthesis*. Oxford: Blackwell Science, 2002.

[119] P. Huo and D. F. Coker, "Theoretical study of coherent excitation energy transfer in cryptophyte phycocyanin 645 at physiological temperature," *The Journal of Physical Chemistry Letters*, no. 7, pp. 825–833, 2011.

[120] H. Lee *et al.*, "Coherence dynamics in photosynthesis: Protein protection of excitonic coherence," *Science*, vol. 316, pp. 1462–1465, 2007.

[121] T. Calhoun *et al.*, "Quantum coherence enabled determination of the energy landscape in light-harvesting complex ii," *Journal of Physical Chemistry B*, vol. 113, pp. 16 291–16 295, 2009.

[122] D. Abramavicius *et al.*, "Unravelling coherent dynamics and energy dissipation in photosynthetic complexes by 2d spectroscopy," *Biophysical Journal*, vol. 94, p. 3613–3619, 2008.

[123] P. Rebentrost *et al.*, "Environment-assisted quantum transport," *New Journal of Physics*, vol. 11, p. 033003, 2009.

[124] R. Berera *et al.*, "Ultrafast transient absorption spectroscopy: principles and application to photosynthetic systems," *Photosynthesis Research*, vol. 101, p. 105–118, 2009.

[125] T. Brixner *et al.*, "Phase-stabilized two- dimensional electronic spectroscopy," *Journal of Chemical Physics*, vol. 121, pp. 4221–4236, 2004.

[126] E. Collini and G. Scholes, "Coherent intrachain energy migration in a conjugated polymer at room temperature," *Science*, vol. 323, pp. 369–373, 2009.

[127] A. Olaya-Castro *et al.*, "Efficiency of energy transfer in a light-harvesting system under quantum coherence," *Physical Review B*, vol. 78, p. 085115, 2008.

[128] R. Sension, "Quantum path to photosynthesis," *Nature*, vol. 446, p. 740, 2007.

[129] M. Plenio and S. Huelga, "Entangled light from white noise," *Physical Review Letters*, vol. 88, p. 197901, 2002.

[130] S. Huelga and M. Plenio, "Stochastic resonance phenomena in quantum many-body systems," *Physical Review Letters*, vol. 98, p. 170601, 2007.

[131] Y. Suzuki *et al.*, "Comparative study on model parameter evaluations for the energy transfer dynamics in Fenna-Matthews-Olson complex," *Chemical Physics*, vol. 539, p. 110903, 2020.

[132] L. Banchi *et al.*, "Modelling non-markovian quantum processes with recurrent neural networks," *New Journal of Physics*, vol. 20, p. 123030, 2018.

[133] J. Strümpfer and K. Schulten, "Excited state dynamics in photosynthetic reaction center and light harvesting complex 1," *The Journal of Chemical Physics*, vol. 137, p. 065101, 2012.

[134] A. Ishizaki and Y. Tanimura, "Quantum dynamics of system strongly coupled to low-temperature colored noise bath: reduced hierarchy equations approach," *Journal of the Physical Society of Japan*, vol. 74, p. 3131–3134, 2005.

[135] J. Jin *et al.*, "Exact dynamics of dissipative electronic systems and quantum transport: Hierarchical equations of motion approach," *Journal of Chemical Physics*, vol. 128, p. 234703, 2008.

[136] J. Strümpfer and K. Schulten, "Open quantum dynamics calculations with the hierarchy equations of motion on parallel computers," *Journal of Chemical Theory and Computation*, vol. 8, p. 2808, 2012.

[137] H. Liu *et al.*, "Reduced quantum dynamics with arbitrary bath spectral densities: Hierarchical equations of motion based on several different bath decomposition schemes," *Journal of Chemical Physics*, vol. 140, p. 134106, 2014.

[138] A. G. Redfield, "On the theory of relaxation processes," *IBM Journal of Research and Development*, vol. 1, pp. 19–31, 1957.

[139] W. T. Pollard *et al.*, "The redfield equation in condensed-phase quantum dynamics," *Advances in Chemical Physics*, vol. 93, pp. 77–134, 1996.

[140] T. Förster, "Zwischenmolekulare energiewanderung und fluoreszenz," *Annals of Physics.*, vol. 437, pp. 55–75, 1948.

[141] T. Förster *et al.*   New York, Academic Press, 1965, pp. 93–137.

[142] A. Ishizaki and G. R. Fleming, "Theoretical examination of quantum coherence in a photosynthetic system at physiological temperature," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, pp. 17 255–17 260, 2009.

[143] A. Ishizaki *et al.*, "Quantum coherence and its interplay with protein environments in photosynthetic electronic energy transfer," *Physical Chemistry Chemical Physics*, vol. 12, pp. 7319–7337, 2010.

[144] J. M. Stokes *et al.*, "A deep learning approach to antibiotic discovery," *Cell*, vol. 180, p. 688–702, 2020.

[145] R. Gómez-Bombarelli *et al.*, "Design of efficient molecular organic light-emitting diodes by a high-throughput virtual screening and experimental approach," *Nature Materials*, vol. 15, p. 1120, 2016.

[146] M. Sumita *et al.*, "Hunting for organic molecules with artificial intelligence: molecules optimized for desired excitation energies," *ACS Central Science*, vol. 4, p. 1126–1133, 2018.

[147] P. Domingos, "A few useful things to know about machine learning," *Communications of the ACM*, vol. 55, pp. 78–87, 2012.

[148] K. P. Bennett and E. Parrado-Hernández, "The interplay of optimization and machine learning research," *Journal of Machine Learning Research*, vol. 7, p. 1265–1281, 2006.

[149] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *Bulletin of Mathematical Physics*, vol. 5, p. 115, 1943.

[150] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *Proceedings of the National Academy of Sciences*, vol. 79, p. 2554–2558, 1982.

[151] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain," *Psychological Review*, vol. 65, pp. 386–408, 1958.

[152] D. E. Rumelhart *et al.*, "Learning representations by back-propagating errors," *Nature*, vol. 323, p. 533–536, 1986.

[153] D. E. Rumelhart, G. E. Hinton *et al.*, "Learning internal representations by error propagation," *Technical report, DTIC Document*, 1985.

[154] P. Werbos, "Beyond regression : new tools for prediction and analysis in the behavioral sciences," Ph.D. dissertation, Harvard University, 1975.

[155] Y. LeCun *et al.*, "Backpropagation applied to handwritten zip code recognition," *Neural Computation*, vol. 1, pp. 541–551, 1989.

[156] G. Hinton *et al.*, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, p. 1527–1554, 2006.

[157] Z. Meng *et al.*, "Using a data driven approach to predict waves generated by gravity driven mass flows," *Water*, vol. 12, p. 600, 2020.

[158] M. Minsky and S. Papertd, *Perceptrons: An introduction to computational geometry*. Cambridge: MIT Press, 1969.

[159] K. Naicker *et al.*, "Machine learning for excitation energy transfer dynamics," *Physical Review Research*, vol. 4, p. 033175, Sep 2022. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevResearch.4.033175

[160] G. R. Fleming *et al.*, "Two dimensional electronic spectroscopy of molecular complexes," *Journal of the Chinese Chemical Society*, vol. 53, pp. 15–24, 2006.

[161] G. S. Engel, "Quantum coherence in photosynthesis," *Procedia Chemistry*, vol. 3, pp. 222–231, 2011.

[162] F. D. Fuller *et al.*, "Vibronic coherence in oxygenic photosynthesis," *Nature Chemistry*, vol. 6, pp. 706–711, 2014.

[163] J. Wen *et al.*, "Native electrospray mass spectrometry reveals the nature and stoichiometry of pigments in the FMO photosynthetic antenna protein," *Biochemical Journal*, vol. 50, pp. 3502–3511, 2011.

[164] S. I. E. Vulto *et al.*, "Excited state dynamics in FMO antenna complexes from photosynthetic green sulfur bacteria: A kinetic model," *The Journal of Physical Chemistry B*, vol. 103, pp. 8153–8161, 1999.

[165] H. van Amerongen L. Valkunas and R. van Grondelle, *Photosynthetic Excitons*. World Scientific, Singapore, 2000.

[166] D. M. Wilkins and N. S. Dattani, "Why quantum coherence is not important in the Fenna-Matthews-Olson complex," *Journal of Chemical Theory and Computation*, vol. 11, pp. 3411–3419, 2015.

[167] N. S. Dattani *et al.*, "Optimal representation of the bath response function & fast calculation of influence functional coefficients in open quantum systems with bathfit 1," *arXiv:1205.4651v2*, 2012.

[168] H. G. Duan *et al.*, "Nature does not rely on long-lived electronic quantum coherence for photosynthetic energy transfer," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 114, pp. 8493–8498, 2017.

[169] E. Thyrhaug *et al.*, "Identification and characterization of diverse coherences in the Fenna–Matthews–Olson complex," *Nature Chemistry*, vol. 10, pp. 780–786, 2018.

[170] J. Carrasquilla and R. Melko, "Machine learning phases of matter," *Nature Physics*, vol. 13, pp. 431–434, 2017.

[171] G. Carleo and M. Troyer, "Solving the quantum many-body problem with artificial neural networks," *Science*, vol. 355, pp. 602–606, 2017.

[172] F. Häse *et al.*, "Machine learning for quantum dynamics: deep learning of excitation energy transfer properties," *Journal of Chemical Sciences*, vol. 8, pp. 8419–8426, 2017.

[173] R. A. Vargas-Hernández *et al.*, "Fully differentiable optimization protocols for non-equilibrium steady states," *New Journal of Physics*, vol. 23, p. 123006, 2021.

[174] B. Hou and R. V. Krems, "Quantum transfer through small networks coupled to phonons: Effects of topology versus phonons," *Physical Review E*, vol. 104, p. 045302, 2021.

[175] R. A. Vargas-Hernández *et al.*, "Multi-objective optimization for retinal photoisomerization models with respect to experimental observables," *The Journal of Chemical Physics*, vol. 155, p. 234109, 2021.

[176] K. Miettinen, *Nonlinear Multiobjective Optimization*. Springer US, 1998.

[177] G. Montavon *et al.*, "Machine learning of molecular electronic properties in chemical compound space," *New Journal of Physics*, vol. 15, p. 095003, 2013.

[178] J. Gomes *et al.*, "Atomic convolutional networks for predicting protein-ligand binding affinity," *arXiv:1703.10603*, 2017.

[179] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition*. O'Reilly Media, Inc., 2019.

[180] K. P. Murphy, *Machine Learning: A probabilistic perspective*. MIT Press, 2012.

[181] Y. LeCun *et al.*, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE 86*, vol. 86, pp. 2278 – 2324, 1998.

[182] D. H. Hubel and T. N. Wiesel, "Receptive fields of single neurones in the cat's striate cortex," *The Journal of Physiology*, vol. 148, pp. 574–591, 1959.

[183] K. Fukushima, "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," *Biological Cybernetics*, vol. 36, pp. 193–202, 1980.

[184] I. Goodfellow *et al.*, *Deep Learning*. MIT Press, 2016, http://www.deeplearningbook.org.

[185] S. J. Jang and B. Mennucci, "Delocalized excitons in natural light-harvesting complexes," *Reviews of Modern Physics*, vol. 90, p. 035003, 2018.

[186] C. M. Bishop, *Pattern recognition and machine learning*. Berlin, Heidelberg: Springer, 2006.

[187] J. Nelson *et al.*, "Data-Driven Time Propagation of Quantum Systems with Neural Networks," *arXiv*, 2022.

[188] A. Ullah and P. O. Dral, "Predicting the future of excitation energy transfer in light-harvesting complex with artificial intelligence-based quantum dynamics," *ChemRxiv*, 2022.

[189] L. E. H. Rodriguez *et al.*, "A comparative study of different machine learning methods for dissipative quantum dynamics," *Machine Learning: Science and Technology*, vol. 3, p. 045016, 2022.

[190] J. H. Cochrane, *Time Series for Macroeconomics and Finance*. Graduate School of Business, University of Chicago, 1997.

[191] M. Cho *et al.*, "Exciton analysis in 2d electronic spectroscopy," *The Journal of Physical Chemistry B*, vol. 109, p. 10542, 2005.

[192] G. Zhang, "Time series forecasting using a hybrid arima and neural network model," *Neurocomputing*, vol. 50, p. 159–175, 2003.

[193] G. P. Zhang, "A neural network ensemble method with jittered training data for time series forecasting," *Information Sciences*, vol. 177, p. 5329–5346, 2007.

[194] Y. LeCun *et al.*, "Deep learning," *Nature*, vol. 521, p. 436–444, 2015.

[195] "ETNA," https://github.com/tinkoff-ai/etna, 2022.

[196] G. E. Box *et al.*, *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.

[197] C. Sammut and G. I. Webb, Eds., *Decision Trees For Regression*. Springer, 2010, p. 267.

[198] L. Prokhorenkova *et al.*, "Catboost: unbiased boosting with categorical features," *arXiv*, 2019.

[199] B. Letham and S. Taylor, "Forecasting at scale," *https://doi.org/10.7287/peerj.preprints.3190v2*, 2017.

[200] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[201] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015, software available from tensorflow.org. [Online]. Available: https://www.tensorflow.org/

[202] S. Hochreiter, "Untersuchungen zu dynamischen neuronalen netzen," Master's thesis, Technical University Munich, Institute of Computer Science, 1991.

[203] Y. Bengio *et al.*, "Learning long-term dependencies with gradient descent is difficult," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 5, p. 157, 1994.

[204] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, p. 1735, 1997.

[205] C. Olah, "Introduction to LSTMs with tensorflow," http://colah.github.io, 2015.

[206] S. Bandyopadhyay *et al.*, "Applications of neural networks to the simulation of dynamics of open quantum systems," *Journal of Chemical Physics*, vol. 515, p. 272, 2018.

[207] P. Lara-Benìtez *et al.*, "An experimental review on deep learning architectures for time series forecasting," *International Journal of Neural Systems*, vol. 31, p. 2130001, 2021.

[208] B. Lim and S. Zohren, "Time-series forecasting with deep learning: A survey," *Philosophical Transactions of the Royal Society A*, vol. 379, pp. 20 200–209, 2021.

[209] O. B. Sezer *et al.*, "Financial time series forecasting with deep learning : A systematic literature review: 2005–2019," *Applied Soft Computing*, vol. 90, pp. 106–181, 2020.

[210] H. I. Fawaz *et al.*, "Deep learning for time series classification: A review," *Data mining and knowledge discovery*, vol. 33, p. 917– 963, 2019.

[211] K. Greff *et al.*, "LSTM: A search space odyssey," *arXiv*, 2015.

[212] Y. LeCun *et al.*, "Convolutional networks for images, speech, and time series," *The handbook of brain theory and neural networks*, vol. 3361, 1995.

[213] S. Chibani *et al.*, "Machine learning approaches for the prediction of materials properties," *APL Materials*, vol. 8, p. 080701, 2020.

[214] M. Ceriotti *et al.*, "Machine learning meets chemical physics," *The Journal of Chemical Physics*, vol. 154, p. 160401, 2021.

[215] L. E. H. Rodríguez and A. A. Kananenka, "Convolutional neural networks for long time dissipative quantum dynamics," *The Journal of Physical Chemistry Letters*, vol. 12, p. 2476–2483, 2021.

[216] M. Raissi, P. Perdikaris, and G. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *Journal of Computational Physics*, vol. 378, p. 686, 2019.

[217] A. D. Jagtap and G. E. Karniadakis, "Extended physics-informed neural networks (xpinns): A generalized space-time domain decomposition based deep learning framework for nonlinear partial differential equations," *Communications in Computational Physics*, vol. 28, p. 2002, 2020.

[218] A. D. Jagtap *et al.*, "Physics-informed neural networks for inverse problems in supersonic flows," *Journal of Computational Physics*, vol. 466, p. 111402, 2022.

Appendix

The architecture of the convolutional neural network model developed in Section 6 is summarized in 13 below. Here the number of input features is three and the number of outputs is two. This model can be adapted for systems of varying complexity as seen in this work. The number of input features for an N-pigment system is $2(N) - 1$ and the number of outputs is $2(N - 1)$.
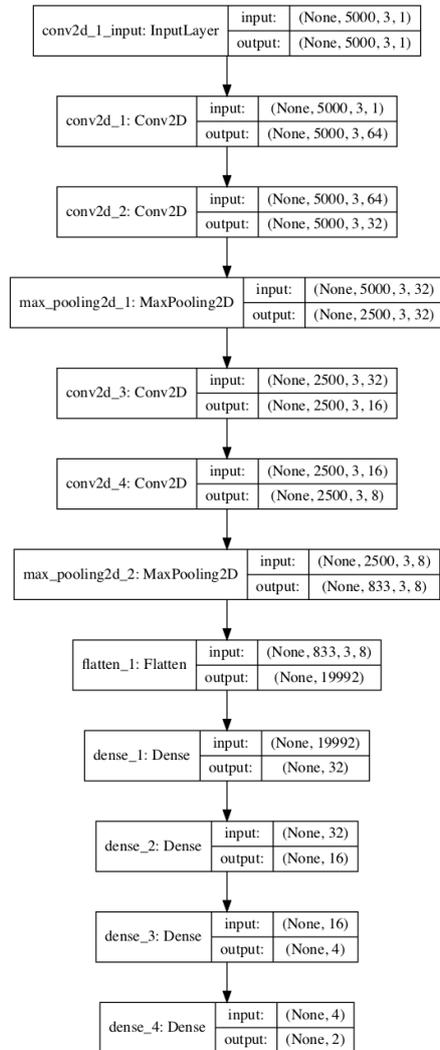
Fig. 13: In this figure, the architecture of the convolutional neural network suited for the prediction of parameters of two-level system Hamiltonians is shown. The input/ output parameters on the right hand side of the figure describe the dimensions of the data going into and out of each layer. The left hand side describes the type of layer used in the network where the following key can be used - conv2d: 2D convolution layer, max_pooling2d: max pooling layer for 2D inputs, flatten: layer to flatten the input and dense: deeply connected layer.