ABSTRACT

Title of Dissertation: CONTROL AND CHARACTERIZATION OF OPEN QUANTUM SYSTEMS

Seyed Alireza Seif Tabrizi
Doctor of Philosophy, 2020

Dissertation Directed by: Professor Mohammad Hafezi
Department of Physics

The study of open physical systems concerns finding ways to incorporate the lack of information about the environment into a theory that best describes the behavior of the system. We consider characterizing the environment by using the system as a sensor, mitigating errors, and learning the physics governing systems out of equilibrium with computer algorithms.

We characterize long-range correlated errors and crosstalk, which are important factors that negatively impacts the performance of noisy intermediate-scale quantum (NISQ) computing devices. We propose a compressed sensing method for detecting correlated dephasing errors, assuming only that the correlations are sparse (i.e., at most s pairs of qubits have correlated errors, where $s \ll n(n-1)/2$, and n is the total number of qubits). Our method uses entangled many-qubit GHZ states, and it can detect long-range correlations whose distribution is completely arbitrary, independent of the geometry of the system. Our method is also highly scalable: it requires only $s \log n$ measurement settings, in contrast to the naive $O(n^2)$ estimate, and efficient classical postprocessing based on convex optimization.

For mitigating the effect of errors, we consider measurements in a quantum computer. We exploit a simple yet versatile neural network to classify multi-qubit quantum states, which is trained using experimental data. We experimentally illustrate this approach in the readout of trapped-ion qubits using additional spatial and temporal features in the data. Using this neural network classifier, we efficiently treat qubit readout crosstalk, resulting in a 30% improvement in detection error over the conventional threshold method. Our approach does not depend on the specific details of the system and can be readily generalized to other quantum computing platforms.

To learn about physical systems using computer algorithms, we consider the problem of arrow of time. We show that a machine learning algorithm can learn to discern the direction of time's arrow when provided with a system's microscopic trajectory as input. Examination of the algorithm's decision-making process reveals that it discovers the underlying thermodynamic mechanism and the relevant physical observables. Our results indicate that machine learning techniques can be used to study systems out of equilibrium, and ultimately to uncover physical principles.

CONTROL AND CHARACTERIZATION OF OPEN QUANTUM SYSTEMS

by

Seyed Alireza Seif Tabrizi

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2020

Advisory Committee:
Professor Mohammad Hafezi, Chair/Advisor
Professor Zohreh Davoudi
Professor Christopher Jarzynski
Professor Norbert M. Linke
Professor Yi-Kai Liu

Acknowledgments

My journey through graduate school and the research that I have done would not have been possible without the help and support of many people that I would like to thank.

I would like to thank my family and my parents. They recognized the importance of education and made sure that I receive the best education from my childhood. I would also like to thank my teachers from elementary school to university, to whom I am forever indebted.

I was very fortunate to have Mohammad Hafezi as my advisor, who guided and supported me through every aspect of my life in graduate school. I also had the opportunity to collaborate with and learn from many scientists both at the University of Maryland and other institutions, including the members of my dissertation committee, Norbert Linke, Christopher Jarzynski, Yi-Kai Liu, and Zohreh Davoudi. Each taught me valuable lessons about science and also on being a good scientist. I also learned a great deal from collaborating with Jiehang Zhang, Keivan Esfarjani, Christopher Monroe, Alexey Gorshkov, Mohammad Maghrebi, Peter Zoller, Hannes Pichler, Pedram Roushan, and Paraj Titum.

I also benefited from collaborating with and learning from student and postdocs at the University of Maryland including Wade DeGottardi, Guanyu Zhu, Sunil

Mittal, Hirokazu Miake, Amir Kalev, Mathias Van Regemortel, Yaroslav Kharkov, Hossein Dehghani, Przemyslaw Bienias, Caroline Figgat, Khan Mahmoud, Ze-Pei Cian, Enjui Kuo, Yijia Xu, Amitava Banerjee, Daiwei Zhu, Andrew Shaw, Patrick Becker, Andrew Guo, Eddie Shoute, Shahriar Aghayi, Ali Lavasani, Yahya Alavi, Ali Izadi Rad, and all the other friends whom I may have missed. Finally, I enjoyed the company of my officemates, Zach Eldredge, Hwanmun Kim, and Sabyasachi Barik.

I was privileged to do an internship at IonQ and learn a lot about quantum computing research in the industry. For that, I would like to thank my friends and colleagues at IonQ including Jungsang Kim,, Yunseong Nam, Ken Wright, Jonathan Mizrahi, Kevin Landsman, Ming Li , Neal Pisenti, Kristi Beck, Jason Amini, Jwo-Sy Chen, Shantanu Debnath, Pradeep Niroula, and Niko Grzesiak.

Moving to the United States for my graduate studies was a significant transition. I would like to thank my housemates at Physics House, George Hine, Kiersten Ruisard, Rachel Ruisard, and Madison Anderson for making this transition as smooth as possible. I would also like to thank the extend Physics House: Nat Steinstultz, Meredith Lukow, Joseph Gareth, and Virginia Kotzias.

Finally, I want to thank my wife Maya Amouzegar, whom I also met at the University of Maryland, for her unconditional love and encouragement. I could not imagine finishing this journey without her support.

Table of Contents

Acknowledgements ii

Table of Contents iv

List of Tables vi

List of Figures vii

Citations to Previously Published Work viii

Chapter 1: Introduction 1
- 1.1 Motivation . . . 1
- 1.2 Open quantum systems . . . 2
- 1.3 Dephasing . . . 4
- 1.4 Quantum detectors . . . 7
- 1.5 Steady state dynamics . . . 8
- 1.6 Challenges . . . 9
- 1.7 Outline of Thesis . . . 11

Chapter 2: Compressed sensing measurement of long-range correlated noise 14
- 2.1 Correlated Dephasing . . . 14
- 2.2 Sparsity and Compressed Sensing . . . 14
  - 2.2.1 Dephasing of GHZ-like states . . . 15
  - 2.2.2 Estimating the decay rate $\Gamma_{\mathbf{ab}}$ . . . 17
  - 2.2.3 Reconstructing the correlation matrix C . . . 18
  - 2.2.4 Recovery guarantee . . . 23
  - 2.2.5 Complex C . . . 27
- 2.3 Numerical examples . . . 34
  - 2.3.1 Effect of gate errors . . . 36
- 2.4 Conclusion and outlook . . . 42

Chapter 3: Improving readout fidelity of trapped-ion quantum computers with machine learning 44
- 3.1 Introduction . . . 44
- 3.2 Initialization and Readout . . . 46
- 3.3 Methods . . . 49
- 3.4 Results . . . 52

3.5 Conclusion and outlook . . . 59

Chapter 4: Machine learning thermodynamic arrow of time 61
4.1 Introduction . . . 61
4.2 Thermodynamics and the arrow of time . . . 63
4.3 Neural networks . . . 67
4.4 Case studies . . . 72
4.4.1 Brownian particle in a moving potential . . . 72
4.4.2 Spin chain in a magnetic field . . . 73
4.5 Interpretation and extensions . . . 78
4.5.1 Inceptionism and dreaming . . . 78
4.5.2 Coarse-grained features and the reduced phase-space . . . 78
4.5.3 Mixture of experts . . . 81
4.5.4 Alternative activation functions . . . 83
4.6 Optimal Networks . . . 85
4.6.1 Brownian particle in a moving potential . . . 86
4.6.2 Spin chain - $\mathcal{B}$-protocol . . . 87
4.6.3 Spin chain - $\mathcal{J}$-protocol . . . 87
4.6.4 Coarse-grained features . . . 91
4.7 Conclusion and outlook . . . 94

Chapter 5: Conclusion and Outlook 95

Appendix A: Optimal control for quantum detectors 99
A.1 Summary of the results . . . 99
A.2 Numerical optimization of control pulses . . . 100

Appendix B: Towards analog quantum simulations of lattice gauge theories with trapped ions 107
B.1 Summary of the results . . . 107
B.2 Tuning spin-dependent forces . . . 108
B.2.1 Stark shift . . . 110
B.2.2 Spontaneous emission . . . 111
B.2.3 Spin-dependent force . . . 111

Bibliography 113

List of Tables

4.1 The value of L_2 regularization parameter α for the NNs in this Chapter. 71

List of Figures

2.1 Illustration of the noise model and the protocol. 21
2.2 Scaling of the reconstruction error $\|W^{(opt)} - C\|_\infty$ under various circumstances. 35
2.3 Deviations of the exponential decay for different gate errors. 42

3.1 The setup and readout scheme for the trapped-ion quantum computer. 45
3.2 The histogram of observed photons in state detection. 48
3.3 An artificial neuron and a neural network. 50
3.4 Results of an instance of the readout experiment for different initial states. 53
3.5 Comparison of different methods for state detection for three qubits. . 55
3.6 Comparing neural networks with threshold methods for five-qubit state detection. 56
3.7 Recurrent neural network approach for state detection 58

4.1 Non-equilibrium physics, time's arrow, and machine learning. 65
4.2 Brownian particle in a moving potential. 74
4.3 Spin chain in a time-dependent magnetic field. 76
4.4 Spin chain with a time-dependent coupling. 77
4.5 Interpreting the neural network's inner mechanism. 79
4.6 A highly irreversible version of the $\mathcal{B}$-protocol. 82
4.7 Mixture of experts 84
4.8 The theoretical likelihood function and the output of the neural network with custom activation functions. 85
4.9 Performance of logistic regression in the $\mathcal{J}$-protocol. 91
4.10 Prediction of logistic regression with coarse-grained features. 92

A.1 Optimizing for detection for correlated background noise (Lorentzian spectrum) in the SCA. 103
A.2 Comparison of the numerically optimized protocol (SCA-optimal) with spin lock and CPMG protocols for different values of noise correlation σ_t. 105
A.3 The effect of constraints on the optimal protocol for $J\sigma_t = 0.56$. . . 106

B.1 The level diagram of $^{171}Yb^+$ relevant to the scheme presented in this Appendix. 109

Citations to Previously Published Work

Part of this thesis is based on the published work of the author. Here we present the references that each chapter is based on and also mention other relevant publications of the author that do not appear in this thesis.

- Chapter 2 is based on a manuscript that is in preparation

  Seif, Alireza, Mohammad Hafezi, Yi-Kai Liu. "Compressed Sensing Measurement of Long-Range Correlated Noise." *In preparation* (2020).

  - Appendix A is based on author's contribution to

    Titum, Paraj, Kevin M. Schultz, Alireza Seif, Gregory D. Quiroz, and B. D. Clader. "Optimal control for quantum detectors." arXiv preprint arXiv:2005.05995 (2020).

  - Further work related to measurement of entanglement spectrum of cold atoms is presented in

    Pichler, Hannes, Guanyu Zhu, Alireza Seif, Peter Zoller, and Mohammad Hafezi. "Measurement protocol for the entanglement spectrum of cold atoms." Physical Review X 6, no. 4: 041033 (2016).

  - Further work related to engineering controlled dissipation is presented in

    Cian, Ze-Pei, Guanyu Zhu, Su-Kuan Chu, Alireza Seif, Wade DeGottardi, Liang Jiang, and Mohammad Hafezi. "Photon pair condensation

by engineered dissipation." Physical review letters 123, no. 6: 063602 (2019).

  - Further work related to entropy scaling transition in trajectories of a quantum system under different noise processes has appeared in

    Van Regemortel, Mathias, Ze-Pei Cian, Alireza Seif, Hossein Dehghani, and Mohammad Hafezi. "Entanglement entropy scaling transition under competing monitoring protocols." arXiv preprint arXiv:2008.08619 (2020).

- Chapter 3 is based on

  Seif, Alireza, Kevin A. Landsman, Norbert M. Linke, Caroline Figgatt, C. Monroe, and Mohammad Hafezi. "Machine learning assisted readout of trapped-ion qubits." Journal of Physics B: Atomic, Molecular and Optical Physics 51, no. 17: 174006 (2018).

  - Appendix B is based on author's contribution to

    Davoudi, Zohreh, Mohammad Hafezi, Christopher Monroe, Guido Pagano, Alireza Seif, and Andrew Shaw. "Towards analog quantum simulations of lattice gauge theories with trapped ions." Physical Review Research 2, no. 2: 023015 (2020).

- Chapter 4 is based on

  Seif, Alireza, Mohammad Hafezi, and Christopher Jarzynski. "Machine learning the thermodynamic arrow of time." Nature Physics (2020).

– Further work related to engineering phonon transport is presented in

Seif, Alireza, Wade DeGottardi, Keivan Esfarjani, and Mohammad Hafezi. "Thermal management and non-reciprocal control of phonon flow via optomechanics." Nature communications 9, no. 1 : 1-8 (2018).

Chapter 1: Introduction

1.1 Motivation

Since its inception in the early 20th century, quantum mechanics has significantly changed our lives. For example, understanding the wave nature of electrons has led to the invention of transistors, and understanding the quantum statistics of photons has led to the creation of lasers. However, most of these applications involve controlling a collective degree of freedom in a physical system. Currently, we are observing a second wave of inventions that involve meticulous manipulation of individual quantum degrees of freedom, such as the electronic state of an ion [1]. One application of these experimental advancements is the realization of small quantum computers [2]. In particular, these advances have allowed for the creation of tailored patterns of quantum correlations among many individual quantum bits (qubits) known as entanglement [3]. Such correlations cannot exist classically and harnessing them holds the key to finding advantage in quantum technologies over their classical counterparts. This advantage exist in different applications such as sensing or computation. For example, it has been shown that using entangled states enhances sensitivity of the sensors. Additionally, the famed Shor's algorithm utilizes a quantum system to factor integers more efficiently than currently known

classical algorithms [4].

The sensitivity of quantum systems to external variables is both a curse and a blessing. Uncontrolled interaction with the environment destroys the quantum system's coherence and erases the advantage of these systems over their classical counterparts. At the same time, it makes these devices very good sensors[5]. Moreover, understanding how the environment affect the evolution of a quantum system is an interesting theoretical question in itself. The main theme of this thesis is characterizing the effect of environment on the system, undoing its adverse effects, controlling the system in the presence of noise, and finally understanding fundamental physical laws governing the physical systems out of equilibrium.

1.2 Open quantum systems

The dynamics of a closed quantum system is governed by the Schrodinger equation. The state of the system is represented by a vector $|\psi\rangle$ in Hilbert space $\mathcal{H}$. The dynamics is then given by (setting the Planck's constant $\hbar = 1$) [6]

$$\frac{d|\psi\rangle}{dt} = -iH|\psi\rangle, \tag{1.1}$$

where H, the generator of the dynamics, is a Hermitian operator called the Hamiltonian. Consequently, the time evolution operator $U(t,t') = \mathcal{T}\exp[-i\int_{t'}^{t} H(\tau)d\tau]$ is a unitary operator such that $|\psi(t)\rangle = U(t,t')|\psi(t')\rangle$. Here, $\mathcal{T}$ is the time ordering operator.

We now consider a situation, where our universe is comprised of a system

and an environment, in Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$, respectively. We also focus on finite dimensional Hilbert spaces. The overall dynamics of the universe is still unitary and governed by (1.1). However, the system without the information about the environment is now an open quantum system, and its reduced dynamics is no longer unitary [7]. Specifically, let us assume that the universe at $t = 0$ is in the state $|\Psi(0)\rangle = |\psi_1(0)\rangle \otimes |\psi_2(0)\rangle$. After time t we have $|\Psi(t)\rangle = U(t,0)|\Psi(0)\rangle$. If we do not have any information about the environment this point, we have to express the state of the system as a statistical mixture that reflects this lack of knowledge about the state of the environment. This is achieved by using the density matrix formalism and taking the partial trace of the full state of the universe

$$\rho_1(t) = \mathrm{tr}_2[|\Psi(t)\rangle\langle\Psi(t)|] = \sum_i \langle e_i|\Psi(t)\rangle\langle\Psi(t)|e_i\rangle = \mathcal{E}(t,0)[|\psi_1(0)\rangle\langle\psi_1(0)|] \tag{1.2}$$

where $\{|e_i\rangle\}$ is a complete basis for $\mathcal{H}_2$, and $\rho_1(t)$ is a density matrix, a positive semi-definite matrix with unit trace that encodes our knowledge about the system alone. The evolution of the system alone is not necessarily unitary anymore, and is described by a linear completely positive trace preserving map $\mathcal{E}$, that is called a quantum channel. Unlike the unitary dynamics that is always divisible, quantum channels are not necessarily divisible, that is $\mathcal{E}(t,0)$ is not always equal to $\mathcal{E}(t,t')\mathcal{E}(t',0)$. However, there exists an important class of quantum channels, the set of Markovian channels, that are infinitely divisible. In this case, the dynamics of the system follows the Gorini-Kossakowski-Sudarshan-Lindblad equation $\frac{\partial\rho}{\partial t} = \mathcal{L}(\rho)$[8, 9]. Therefore, we have $\mathcal{E}(t) = e^{\mathcal{L}t}$, where we omitted the initial time

$t = 0$ in the notation. The generator $\mathcal{L}$ is called the Lindbladian and has to be in the form [9]

$$\mathcal{L}(\rho) = -i[H, \rho] + \sum_{\alpha,\beta} c_{\alpha,\beta}(L_\alpha \rho L_\beta^\dagger - \frac{1}{2}\{L_\beta^\dagger L_\alpha, \rho\}), \tag{1.3}$$

where H is the system's Hamiltonian, $\{L_\alpha\}$ are called Lindblad operators, $[A, B] = AB - BA$, and $\{A, B\} = AB + BA$. The matrix $C = (c_{ij})$, that is also known as the Kossakowski matrix, has to be positive-semidefinite so that the map $\mathcal{E}$ is completely positive.

The language of open quantum systems is useful when we consider how noise affects our system. After all, noise essentially stems from our lack of knowledge about the system environment interactions, and how the information about the system is lost in the environment degrees of freedom. For example, dephasing noise studied in Chapter 2 and Appendix A can result from either stochastic magnetic field or interactions with another quantum system. In the former, our lack of knowledge corresponds to the stochasticity of the magnetic field, whereas in the latter the information about the state of the environment is lacking. Another example is the spontaneous emission of atoms, which we discuss in Chapter 3.

1.3 Dephasing

A common noise process in physical systems is dephasing. As the name suggests, it corresponds to the loss of phase information in a quantum state. Under dephasing the off-diagonal elements of the density matrix in the energy basis, also known as cohrences, decay to zero.

As noted earlier, the noise process can either result from a classical stochastic process or from interactions with another quantum system. Here, we study the example of a qubit's time evolution under classical Gaussian dephasing noise[5].

Consider a two-level system that evolves under

$$H = \frac{1}{2}\omega_0\sigma_z + \frac{1}{2}\delta B(t)\sigma_z, \tag{1.4}$$

where ω_0 is the energy splitting between the two levels, σ_z is the Pauli z matrix, and $\delta B(t)$ is a time-dependent fluctuating magnetic field, that is a stationary Gaussian process such that $\mathbb{E}[\delta B(t)] = 0$ and $\mathbb{E}[\delta B(t)\delta B(t')] = K(t-t')$. For simplicity we set $\omega_0 = 0$, which in effect is similar to transforming to a rotating frame with that frequency. In this example, dephasing eventually transforms a superposition state, $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$, to a balanced statistical mixture of $|0\rangle$ and $|1\rangle$ states, where $|0\rangle$ and $|1\rangle$ are eigenstates of σ_z corresponding to eigenvalues ± 1, respectively. To see this more clearly, consider a system intially in the state $\rho(0) = \begin{bmatrix} \rho_{00} & \rho_{01} \\ \rho_{10} & \rho_{11} \end{bmatrix}$ at $t = 0$. It evolves to

$$\rho(t) = \exp[-\frac{i}{2}\int_0^t \delta B(s)\sigma_z ds]\rho(0)\exp[\frac{i}{2}\int_0^t \delta B(s)\sigma_z ds] \tag{1.5}$$

$$= \begin{bmatrix} \rho_{00} & \rho_{01}e^{-i\theta(t)} \\ \rho_{10}e^{i\theta(t)} & \rho_{11} \end{bmatrix}, \tag{1.6}$$

where $\theta(t) = \int_0^t \delta B(s)ds$. Note that $\theta(t)$ is a random variable. Therefore, we have

to average over different realizations of noise to find

$$\bar{\rho}(t) = \mathbb{E}[\rho(t)] = \begin{bmatrix} \rho_{00} & \rho_{01}e^{-\gamma(t)} \\ \rho_{10}e^{-\gamma(t)} & \rho_{11} \end{bmatrix}, \tag{1.7}$$

where $\gamma(t) = \frac{1}{2}\int_0^t \int_0^t \mathbb{E}[\delta B(s)\delta B(s')]dsds'$. We see that even though the additional picked up phase $\theta(t)$ has a well-defined value in each realization, our lack of knowledge about its value translates to the decay of coherences on average. The decay factor $\gamma(t)$ is more commonly expressed as

$$\gamma(t) = \frac{1}{2}\int_{-\infty}^{\infty} \frac{d\omega}{2\pi} W_t(\omega) J(\omega), \tag{1.8}$$

where $W_t(\omega) = \frac{4}{\omega^2}\sin^2(\omega t/2)$ is called the window function, and $J(\omega)$ is the Fourier transform of $K(t)$ and is called the spectral density of noise. In the Markovian limit, $\delta B(t)$ is white noise and we have $K(t-t') = 2\gamma_0\delta(t-t')$. Therefore, we find that $\gamma(t) = \gamma_0 t$, which leads to the familiar exponential decay.

Given the explicit expression for the noise-averaged density matrix $\bar{\rho}(t)$ in the Markovian limit, we can take its derivative and find that

$$\frac{\partial\bar{\rho}}{\partial t} = \frac{\gamma_0}{2}(\sigma_z\bar{\rho}\sigma_z - \bar{\rho}) = \frac{\gamma_0}{2}(\sigma_z\bar{\rho}\sigma_z - \frac{1}{2}\{\sigma_z\sigma_z, \bar{\rho}\}), \tag{1.9}$$

which is in the Lindblad form (1.3).

While we introduced dephasing through a single qubit example, the situation can be more complex for muli-qubit systems and the noise can be correlated. We

investigate such processes in Chapter 2.

1.4 Quantum detectors

As noted earlier, quantum systems are sensitive to external environment. Therefore, they can be used to measure unknown signals. For example, one can measure an unknown magnetic field thorough Ramsey interferometry [10]. Specifically, consider a qubit evolving under the Hamiltonian $H = \frac{\omega_0}{2}\sigma_z + \frac{1}{2}B_0\sigma_z$, where B_0 is an unknown constant magnetic field that we wish to measure. Again, let us set $\omega_0 = 0$ for simplicity. We can measure B_0 as follows:

1. Prepare the qubit in the $|\psi_0\rangle = |0\rangle$ state.
2. Apply $U_{\text{Hadamard}} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ to $|\psi_0\rangle$ to get $|\psi_1\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$.
3. Let the system evolve under H for some time t. The state of the system is $|\psi_2\rangle = \frac{1}{\sqrt{2}}(e^{-iB_0t/2}|0\rangle + e^{iB_0t/2}|1\rangle)$ after this step.
4. Apply $U^{\dagger}_{\text{hadamard}}$ to get $|\psi_3\rangle = \cos(B_0t/2)|0\rangle - \sin(B_0t/2)|1\rangle$.
5. Measure the qubit in the σ_z basis, and obtain $\langle\sigma_z(t)\rangle = \cos^2(B_0t/2) - \sin^2(B_0t/2) = \cos(B_0t)$.

Hence, B_0 can be obtained from $\langle\sigma_z(t)\rangle$, by measuring it at different times. This procedure is known as Ramsey interferometry. This protocol can also be used to characterize dephasing noise in the system. Specifically, in the Markovian case the

oscillatory signal obtained in step 5, becomes an exponanetially decaying signal, and from the decay rate one can obtained γ_0 that was discussed in Sec. 1.3.

Extensions of the concepts discussed above appear in quantum metrology and quantum noise spectroscopy [5]. For example, it has been shown that using entangled states can improve the sensitivity of the measurement device. Another extension is the study of non-Markovian noise, where one needs to apply a control Hamiltonian in order to access information about $J(\omega)$ for different values of ω. Additionally, the noise can be spatially correlated, which requires an extension of Ramsey experiments to multiple qubits to learn about correlations.

Additionally, determining the best control strategy to achieve a given task in the presence of noise has many practical applications. For example, one can perform higher fidelity quantum gates by looking for control strategies that compensate for noise [11, 12, 13].

1.5 Steady state dynamics

An important question in the study of open systems is determining the properties of the steady state of the system, that is the state(s) that system reaches at $t \to \infty$. For example, consider a quantum system evolving under the Lindblad equation (1.3). For many systems, there exist a state or multiple states such that $\mathcal{L}(\rho) = 0$. The symmetries, entanglement properties, and correlations of the steady state as a function of the parameters of the problem has long been a subject of interest in the study of open quantum systems [14, 15, 16].

Even in the classical domain, there are some open questions regarding the steady state of the system. For example, it is not known if there is a principle similar to maximum entropy principle, that gives us a recipe for finding steady state of non-equilibrium systems. This is in contrast to problems in equilibrium physics, where we know that systems always try to minimize their free energy.

In recent decades there has been an increased interest in the out-of-equilibrium physics of microscopic systems, leading to a deepened understanding of non-equilibrium fluctuations and their relation to the second law of thermodynamics [17, 18, 19]. One example of these advancements, is the quantification of the time's arrow. In particular, it has become appreciated that fluctuations lead to an effective "blurring" of time's arrow at the nanoscale, and that our ability to discern its direction can be quantified in a system-independent manner [17, 20, 21].

Motivated by these advancements and in hope of developing techniques for understanding physical laws using computer algorithms, and possibly discovering new ones, we study how machine learning tools can be utilized to analyze this arrow of time problem and to identify the relevant physical observables.

1.6 Challenges

As the size of quantum systems grows, their description become more complex. In fact, resources for completely characterizing quantum systems using quantum state and quantum process tomography scales exponentially with the system size [22].

To overcome the limitation imposed by the aforementioned exponential scaling of resources we may want to use prior knowledge about the system under study or make certain assumptions about the model that is used to describe the system. However, it might be difficult to find an effective model for the process of interest. In these cases, it might be advantageous to use a black box approach.

To contrast these two approaches, consider the task of correcting errors in measuring the state of a qubit in an experiment, which we extensively study later in this dissertation. In the first approach, one first finds a mathematical models of the error processes that occur during the measurements, find the parameters of the model using experimental data, and use that model to correct the errors. However, In the black box approach the whole process is considered to be black box whose input is the measurement signal, and the output is the correct quantum state that the signal corresponds to.

Therefore, to build and operate a large scale quantum systems we need a combination of these techniques to both efficiently characterize the dynamics of the system and mitigate the errors when a complete understanding is not available.

Lastly, there has been a great interest in applying machine learning tools to problems in many areas physics of physics, such as statistical mechanics, particle physics, and quantum theory [23]. An outstanding challenge in using these techniques in the study of physics, and science in general, is to develop algorithms that are interpretable [24]. Many of the powerful machine learning algorithms act as black boxes that perform well on the desire task, but offer no understanding of the underlying process. As mentioned above, in certain cases this lack of understanding

is acceptable. However, for the purpose of scientific discovery and gaining insight about physical processes it is important to develop interpretable tools. The existing theories and available experimental and numerical data can serve as a test bed for building such a tool box with the hope that ultimately it could be used to answer open questions in physics.

1.7 Outline of Thesis

The rest of this thesis is organized follows:

- In Chapter 2 we discuss correlated dephasing. Long-range correlated errors and crosstalk are an important factor that negatively impacts the performance of noisy intermediate-scale quantum (NISQ) computing devices. Characterizing these errors can be particularly difficult when there are many qubits, and when one has limited prior information about which qubits are affected. We propose a compressed sensing method for detecting correlated dephasing errors, assuming only that the correlations are sparse (i.e., at most s pairs of qubits have correlated errors, where $s \ll n(n-1)/2$, and n is the total number of qubits). Our method uses entangled many-qubit GHZ states, and it can detect long-range correlations whose distribution is completely arbitrary, independent of the geometry of the system. Our method is also highly scalable: it requires only $s \log n$ measurement settings, and efficient classical postprocessing based on convex optimization.

  Later in Appendix A we discuss a control problem in a noisy environment.

We identify, under modest assumptions about the statistical properties of the signal and noise, the optimal quantum control to detect an external signal in the presence of background noise using a quantum sensor. This part is based on Ref. [25]

- In Chapter 3 we consider how noise and cross-talks affect the measurement accuracy of a physical system. We use machine learning techniques to reduce measurement errors in a quantum computer . We exploit a simple yet versatile neural network to classify multi-qubit quantum states, which is trained using experimental data. This flexible approach allows the incorporation of any number of features of the data with minimal modifications to the underlying network architecture. We experimentally illustrate this approach in the readout of trapped-ion qubits using additional spatial and temporal features in the data. Using this neural network classifier, we efficiently treat qubit readout crosstalk, resulting in a 30% improvement in detection error over the conventional threshold method. Our approach does not depend on the specific details of the system and can be readily generalized to other quantum computing platforms. The results presented in this Chapter have been previously appeared in Ref. [26].

  Later in Appendix B we discuss how trapped-ion qubits can be utilized to simulate other quantum systems. Specifically, we devise practical proposals for analog simulation of simple lattice gauge theories whose dynamics can be mapped onto spin-spin interactions in any dimension. This part is based on

Ref. [27].

- In Chapter 4 we study classical systems out of equilibrium and use machine learning to learn the direction of time's arrow. We show that a machine learning algorithm can learn to discern the direction of time's arrow when provided with a system's microscopic trajectory as input. The performance of our algorithm matches fundamental bounds predicted by nonequilibrium statistical mechanics. Examination of the algorithm's decision-making process reveals that it discovers the underlying thermodynamic mechanism and the relevant physical observables. Our results indicate that machine learning techniques can be used to study systems out of equilibrium, and ultimately to uncover physical principles. The Chapter previously appeared in Ref. [28].

Chapter 2: Compressed sensing measurement of long-range correlated noise

2.1 Correlated Dephasing

We consider a model of spatially correlated noise, for a system that consists of n qubits. We consider the case of dephasing noise, where $H = 0$, and $L_j = Z_j$ (the Pauli σ_z operator acting on the j'th qubit) in the Lindblad equation (1.3)

$$\frac{d\rho}{dt} = \mathcal{L}(\rho) = \sum_{j,k=1}^{n} c_{jk} \left(Z_k \rho Z_j^\dagger - \frac{1}{2}\{Z_j^\dagger Z_k, \rho\} \right). \tag{2.1}$$

As noted in Chapter 1, the noise is then fully described by the matrix $C = (c_{jk}) \in \mathbb{C}^{n\times n}$. The elements c_{jk} indicate correlations between the noise affecting qubits j and k (see Fig. 2.1 a and b). In order for the time evolution of the system to be a completely positive map, the matrix C must be positive semidefinite.

2.2 Sparsity and Compressed Sensing

Our main contribution is an efficient method for learning the off-diagonal part of the correlation matrix C, under the assumption that it is *sparse*, i.e., it has at

most s nonzero elements, where $s \ll n(n-1)/2$.

For simplicity, we first consider the special case where C is real, and its diagonal elements are known exactly. (Later, we will show how our method can be extended to handle complex C, whose diagonal elements are only known approximately.)

2.2.1 Dephasing of GHZ-like states

We begin by describing a subroutine that allows us to measure certain linear functions of the matrix C. This is essentially a generalization of Ramsey spectroscopy discussed in Chapter 1 (see also Fig. 2.1c). Consider an n-qubit state of the form

$$|\psi_{\mathbf{ab}}\rangle = \frac{1}{\sqrt{2}}(|\mathbf{a}\rangle + |\mathbf{b}\rangle) \in (\mathbb{C}^2)^{\otimes n}, \tag{2.2}$$

where $\mathbf{a}, \mathbf{b} \in \{0,1\}^n$, $|\mathbf{a}\rangle = |a_1, a_2, \ldots, a_n\rangle$ and $|\mathbf{b}\rangle = |b_1, b_2, \ldots, b_n\rangle$. Note that, by choosing $\mathbf{a}$ and $\mathbf{b}$ appropriately, one can make $|\psi_{\mathbf{ab}}\rangle$ be a single-qubit $|+\rangle$ state, a two-qubit Bell state, or a many-qubit GHZ state (while the other qubits are in a tensor product of $|0\rangle$ and $|1\rangle$ states).

Say we prepare the state $|\psi_{\mathbf{ab}}\rangle$, then allow it to evolve for time t under the Lindbladian (2.1), see Fig. 2.1(d). Let $\rho(t)$ be the resulting state. The coherences in the state decay as follows:

$$\rho(t) = \frac{1}{2}\Big(|\mathbf{a}\rangle\langle\mathbf{a}| + e^{-\Gamma_{\mathbf{ab}}t}|\mathbf{a}\rangle\langle\mathbf{b}| + e^{-\Gamma_{\mathbf{ab}}t}|\mathbf{b}\rangle\langle\mathbf{a}| + |\mathbf{b}\rangle\langle\mathbf{b}|\Big) \tag{2.3}$$

where the decay rate $\Gamma_{\mathbf{ab}} \in \mathbb{R}$ is defined so that $\mathcal{L}(|\mathbf{a}\rangle\langle\mathbf{b}|) = -\Gamma_{\mathbf{ab}}|\mathbf{a}\rangle\langle\mathbf{b}|$.

We claim that, by estimating the decay rate $\Gamma_{\mathbf{ab}}$, we can measure a certain linear function of the correlation matrix C. To see this, we write:

$$\Gamma_{\mathbf{ab}} = -\sum_{ij} c_{ij}[(-1)^{a_i+b_j} - \tfrac{1}{2}(-1)^{a_i+a_j} - \tfrac{1}{2}(-1)^{b_i+b_j}] \tag{2.4}$$

$$= -\sum_{ij} c_{ij}[\alpha_i\beta_j - \tfrac{1}{2}\alpha_i\alpha_j - \tfrac{1}{2}\beta_i\beta_j], \tag{2.5}$$

where we defined $\alpha_i = (-1)^{a_i}$ and $\beta_i = (-1)^{b_i}$. We can then express the above equation as

$$\Gamma_{\mathbf{ab}} = -\boldsymbol{\alpha}^\intercal C\boldsymbol{\beta} + \tfrac{1}{2}\boldsymbol{\alpha}^\intercal C\boldsymbol{\alpha} + \tfrac{1}{2}\boldsymbol{\beta}^\intercal C\boldsymbol{\beta} \tag{2.6}$$

$$= -\tfrac{1}{2}[\boldsymbol{\alpha}^\intercal C(\boldsymbol{\beta}-\boldsymbol{\alpha}) + (\boldsymbol{\alpha}-\boldsymbol{\beta})^\intercal C\boldsymbol{\beta}] \tag{2.7}$$

$$= -\tfrac{1}{4}[\boldsymbol{\alpha}^\intercal C(\boldsymbol{\beta}-\boldsymbol{\alpha}) + (\boldsymbol{\alpha}-\boldsymbol{\beta})^\intercal C\boldsymbol{\beta} + \boldsymbol{\beta}^\intercal C(\boldsymbol{\alpha}-\boldsymbol{\beta}) + (\boldsymbol{\beta}-\boldsymbol{\alpha})^\intercal C\boldsymbol{\alpha}] \tag{2.8}$$

$$= \tfrac{1}{2}(\boldsymbol{\alpha}-\boldsymbol{\beta})^\intercal C(\boldsymbol{\alpha}-\boldsymbol{\beta}) \tag{2.9}$$

$$= \tfrac{1}{2}(2\mathbf{r}^\intercal)C(2\mathbf{r}) \tag{2.10}$$

$$= 2\mathbf{r}^\intercal C\mathbf{r}, \tag{2.11}$$

where the symmeterization from the second to the third line was possible because C is real and $C = C^\intercal$, and we also defined a new variable $\mathbf{r} = \frac{\boldsymbol{\alpha}-\boldsymbol{\beta}}{2} \in \{1, 0, -1\}^n$. Note that $\mathbf{r}$ can also be defined in terms of $\mathbf{a}$ and $\mathbf{b}$, via the identity:

$$\mathbf{r} = \mathbf{b} - \mathbf{a}. \tag{2.12}$$

2.2.2 Estimating the decay rate $\Gamma_{\mathbf{ab}}$

There are many possible ways to estimate the decay rate $\Gamma_{\mathbf{ab}}$. For concreteness, we describe one simple method here

1. Choose some evolution time $t \geq 0$ such that $\frac{1}{2} \leq \Gamma_{\mathbf{ab}} t \leq 2$. This ensures that the variance of our estimator for $\Gamma_{\mathbf{ab}}$, which is introduced later, is bounded. We assume that such t can be chosen in advance or adaptively. The detailed analysis of the convergence of errors will be presented in a future work.

2. Repeat the following experiment N_{trials} times:

   - Prepare the state $|\psi_{\mathbf{ab}}\rangle = \frac{1}{\sqrt{2}}(|\mathbf{a}\rangle + |\mathbf{b}\rangle)$, allow the state to evolve for time t, then measure in the basis $\frac{1}{\sqrt{2}}(|\mathbf{a}\rangle \pm |\mathbf{b}\rangle)$.

   Let N_+ and N_- be the number of $\frac{1}{\sqrt{2}}(|\mathbf{a}\rangle + |\mathbf{b}\rangle)$ and $\frac{1}{\sqrt{2}}(|\mathbf{a}\rangle - |\mathbf{b}\rangle)$ outcomes, respectively. Note that the probabilities of these outcomes are given by $P_+ = \frac{1}{2}(1 + e^{-\Gamma_{\mathbf{ab}} t})$ and $P_- = \frac{1}{2}(1 - e^{-\Gamma_{\mathbf{ab}} t})$.

3. Define:

$$\Delta = \frac{N_+ - N_-}{N_{\text{trials}}}. \tag{2.13}$$

   Then return the following estimate for $\Gamma_{\mathbf{ab}}$:

$$\hat{\Gamma}_{\mathbf{ab}} = -\frac{1}{t} \ln(\Delta). \tag{2.14}$$

Note that Δ is an unbiased estimator for $P_+ - P_-$, that is, $\mathbb{E}(\Delta) = P_+ - P_- = e^{-\Gamma_{\mathbf{ab}} t}$. This motivates our definition of the estimator $\hat{\Gamma}_{\mathbf{ab}}$.

2.2.3 Reconstructing the correlation matrix C

The relationship (2.11) suggests a recipe for estimating C. By preparing different states $|\psi\rangle$ of the form (2.2), and measuring the decay rate Γ of the coherences, we can estimate certain linear functions of C, which are proportional to $\mathbf{r}^\intercal C \mathbf{r}$. Here, we propose a method for estimating C efficiently when its off-diagonal elements are sparse. Our method consists of two steps: first, we perform single-qubit spectroscopy in order to learn the diagonal elements of C; second, we apply techniques from compressed sensing (e.g., random measurements, and ℓ_1-minimization) in order to recover the off-diagonal elements of C.

First, we can estimate the diagonal elements c_{jj}, for $j = 1, \ldots, n$, as follows:

1. Let $\mathbf{a} = (0, 0, \ldots, 0)$ and $\mathbf{b} = (0, \ldots, 0, 1, 0, \ldots, 0)$ (where the 1 appears in the j'th position). By equation (2.12), we have that $\mathbf{r} = \mathbf{b} - \mathbf{a} = (0, \ldots, 0, 1, 0, \ldots, 0)$.

2. Prepare the state $|\psi\rangle = \frac{1}{\sqrt{2}}(|\mathbf{a}\rangle + |\mathbf{b}\rangle) = |0\rangle \cdots |0\rangle |+\rangle |0\rangle \cdots |0\rangle$, and construct an estimate g_j of the decay rate $\Gamma_{\mathbf{ab}} = 2\mathbf{r}^T C \mathbf{r} = 2c_{jj}$. Note that this is equivalent to performing single-qubit noise spectroscopy on the j'th qubit.

Let $\mathbf{g} = (g_1, \ldots, g_n) \in \mathbb{R}^n$ be the output of this procedure. We can view $\frac{1}{2}\mathbf{g}$ as an estimate of a "sensing operator" that returns the diagonal elements of the matrix C. Furthermore, we assume we know an upper-bound ϵ_1 on the error of this estimate

(in ℓ_2 norm, say). Formally, we write this as

$$\|\tfrac{1}{2}\mathbf{g} - \mathrm{diag}(C)\|_2 \leq \epsilon_1, \tag{2.15}$$

and we define the linear operator diag : $\mathbb{R}^{n\times n} \to \mathbb{R}^n$ where $\mathrm{diag}(C) = (c_{11}, c_{22}, \ldots, c_{nn})$.

In order to estimate the off-diagonal elements of C, we will use a certain type of random measurement. First, we fix some probability $p \in [0, \frac{1}{2}]$, and we choose some parameter m, which controls the number of different measurements. (We will specify the choice of m later.) Now, for $j = 1, \ldots, m$, perform the following procedure:

1. Choose vector $\mathbf{a}, \mathbf{b} \in \{0,1\}^n$ at random, by choosing each coordinate a_i and b_i independently at random with the following probabilities:

$$a_i, b_i = \begin{cases} 1 & \text{with probability } p, \\ 0 & \text{with probability } 1-p. \end{cases} \tag{2.16}$$

2. Define $\mathbf{r} = \mathbf{b} - \mathbf{a} \in \{-1, 0, 1\}^n$ as given by equation (2.12). Note that $r_i = -1, 1$ with probability $\tilde{p} = p(1-p)$, and $r_i = 0$ with probability $1 - \tilde{p} = p^2 + (1-p)^2$.

3. Prepare the state $|\psi\rangle$ as follows, identical to equation (2.2):

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|\mathbf{a}\rangle + |\mathbf{b}\rangle). \tag{2.17}$$

This is a GHZ state on a subset of the qubits, with some bit flips. It can be

obtained by preparing those qubits i with $a_i = b_i$ in the state $|a_i\rangle$, preparing qubits with $a_i \neq b_i$ in a GHZ state, and applying a Pauli X gate on those qubits with $a_i = 1$. Let $U_{\mathbf{ab}}$ denote the overall operation that prepares this state.

4. Construct an estimate h_j of the decay rate $\Gamma_{\mathbf{ab}} = 2\mathbf{r}^T C \mathbf{r}$.

Let $\mathbf{h} = (h_1, \ldots, h_m) \in \mathbb{R}^m$ be the output of the above procedure. Again, we can view $\frac{1}{2}\mathbf{h}$ as an estimate of a "sensing operator" Φ that is applied to the matrix C. We assume we know an upper-bound ϵ_2 on the error of this estimate (in ℓ_2 norm). Formally, we write

$$\|\tfrac{1}{2}\mathbf{h} - \Phi(C)\|_2 \leq \epsilon_2, \tag{2.18}$$

where we let $\mathbf{r}^{(1)}, \mathbf{r}^{(2)}, \ldots, \mathbf{r}^{(m)}$ be independent random vectors chosen from the same distribution as $\mathbf{r}$ (described above), and we define the operator $\Phi : \mathbb{R}^{n \times n} \to \mathbb{R}^m$ such that

$$\Phi(C) = \big((\mathbf{r}^{(1)})^T C \mathbf{r}^{(1)}, (\mathbf{r}^{(2)})^T C \mathbf{r}^{(2)}, \ldots, (\mathbf{r}^{(m)})^T C \mathbf{r}^{(m)}\big). \tag{2.19}$$

In order to recover the off-diagonal elements of C, we solve the following optimization problem. We are given $\mathbf{g}$ and $\mathbf{h}$, and we are promised that the matrix C is positive semidefinite, and its off-diagonal part is sparse. In general, this sparsity constraint leads to an optimization problem that is computationally intractable. However, we will use a well-known strategy from compressed sensing: we will solve a convex relaxation of the above problem, where we minimize the ℓ_1 (vector) norm of the off-diagonal part of the matrix. This can be written as the following convex

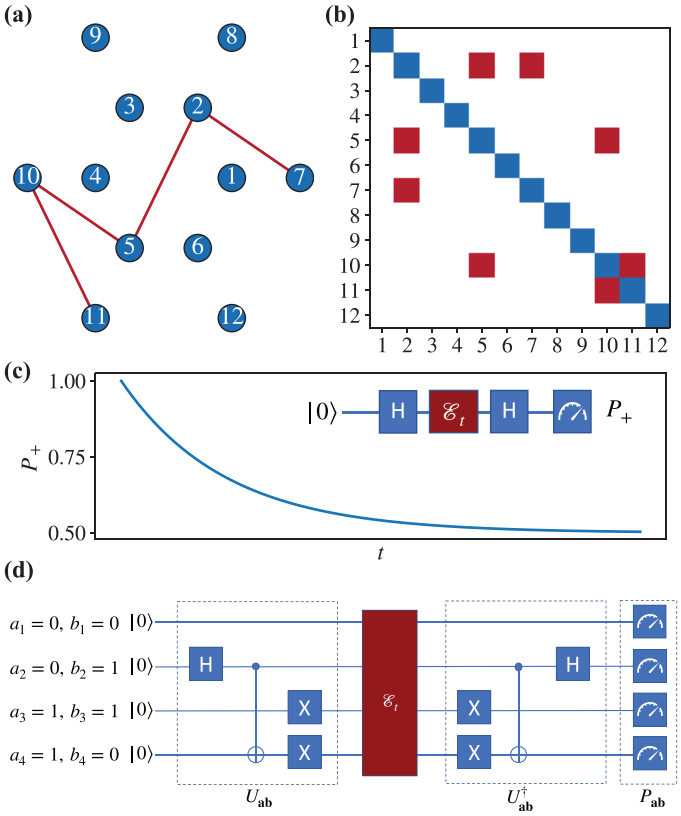

Figure 2.1: Illustration of the noise model and the protocol. (a) The qubits experience correlated Markovian dephasing. The red lines corresponds to non-zero c_{ij}, indicating correlated noise affecting the pairs of qubits connected by those lines.(b) The C matrix corresponding to the correlation graph in panel a. The diagonal elements correspond to single qubit dephasing whereas the off-diagonal elements indicate correlated dephasing noise. (c) Ramsey spectroscopy for characterizing single qubit dephasing. The plot shows the exponential decay of the overlap P_+ with the rate Γ as a function of t. The inset shows the Ramsey protocol, where a superposition of qubit states is prepared with the first Hadamard gate H, the system goes through the dephasing noise channel $\mathcal{E}_t$, and the second H followed by a measurement in the computational basis measures the overlap P_+. (d) The generalized measurement protocol involves generating vectors $\mathbf{a}$ and $\mathbf{b}$ whose elements are randomly chosen from $\{0, 1\}$. We first prepare those qubits i with $a_i = b_i$ in the state $|a_i\rangle$, prepare qubits with $a_i \neq b_i$ in a GHZ state, and apply a Pauli X gate on those qubits with $a_i = 1$. The overall effect of these operations is given by $U_{\mathbf{ab}}$. We then let the system evolves under the dephasing noise for time t, represented by $\mathcal{E}_t$. Finally, we apply $U_{\mathbf{ab}}^\dagger$ and measure the control qubit in the computational basis. The probability of obtaining the outcome 0, gives the overlap $P_{\mathbf{ab}}$. The overlap decays exponentially in time, and we estimate its decay rate to find $\Gamma_{\mathbf{ab}}$. By measuring this decay rate for various $\mathbf{a}$'s and $\mathbf{b}$'s we can recover C.

program:

$$\text{Find } W \in \mathbb{R}^{n\times n} \text{ that minimizes } \sum_{i\neq j} |w_{ij}|, \text{ such that:} \tag{2.20}$$

$$\text{diag}(W) = \tfrac{1}{2}\mathbf{g}, \tag{2.21}$$

$$\|\tfrac{1}{2}\mathbf{h} - \Phi(W)\|_2 \leq \epsilon_2 + \sqrt{m}\sqrt{n}\epsilon_1, \tag{2.22}$$

$$W \succeq 0. \tag{2.23}$$

Note that we have constructed the constraints (2.21) and (2.22) slightly differently from the error bounds (2.15) and (2.18) that we were given. In particular, in (2.21) we force the diagonal elements of W to be equal to $\frac{1}{2}\mathbf{g}$, even though $\frac{1}{2}\mathbf{g}$ may contain an error of size ϵ_1. To allow for this additional error in W, we add $\|\Phi(\text{diag}(\text{diag}(C) - \frac{1}{2}\mathbf{g}))\| \leq \sqrt{m}\sqrt{n}\epsilon_1$ on the right-hand side of (2.22) and use the fact that Φ is acting on a matrix that is diagonal. Constructing the constraints this way makes the convex program simpler to analyze, since it removes the optimization over the diagonal elements of W.

Nonetheless, it is natural to ask what happens if we replace (2.21) and (2.22) with the more obvious constraints:

$$\|\tfrac{1}{2}\mathbf{g} - \text{diag}(W)\|_2 \leq \epsilon_1, \tag{2.24}$$

$$\|\tfrac{1}{2}\mathbf{h} - \Phi(W)\|_2 \leq \epsilon_2. \tag{2.25}$$

We expect that this will not make much of a difference, for most experimental setups. In most experiments, single-qubit spectroscopy is fairly easy to perform, hence we

can estimate $\mathbf{g}$ quite accurately, so we are in the regime where $\epsilon_1 \ll \epsilon_2$. In this regime, the preceding two sets constraints are approximately the same.

Since this is a convex program, it can be solved efficiently (both in theory and in practice), for instance by using interior point algorithms. Let $W^{(opt)}$ be the optimal solution. We want to show that $W^{(opt)} \approx C$. In the following section, we will prove a rigorous recovery guarantee that upper-bounds the error $\|W^{(opt)} - C\|_F$ (in the Frobenius norm, which is equivalent to the 2-norm for vectors).

2.2.4 Recovery guarantee

We now want to analyze the solution $W^{(opt)}$ to the convex program in (2.20)-(2.23), and upper-bound the recovery error $\|W^{(opt)} - C\|_F$. To do this, we will further relax the convex program, by removing the positivity constraint (2.23). This makes the program simpler to analyze, and in fact, we will be able to prove a strong recovery guarantee even without using the positivity constraint (2.23).

(In practice, we would of course solve the convex program with the positivity constraint, in order to get a physically plausible result. Alternatively, we could solve the convex program without the positivity constraint, and then project the solution onto the positive semidefinite cone. We expect that enforcing positivity will typically improve the accuracy of the solution.)

We first state the main result of this chapter informally and prove it in the following.

Let C denote the correlation matrix of the noise, with s non-zero off-diagonal

elements. Assume that we have an estimate $\frac{1}{2}\mathbf{g}$ for the diagonal elements of C that satisfies $\|\frac{1}{2}\mathbf{g} - \text{diag}(C)\|_2 \leq \epsilon_1$. Additionally, assume that we have a noisy estimate $\frac{1}{2}\mathbf{h}$ for the sensing operator $\Phi(C)$ that satisfies $\|\frac{1}{2}\mathbf{h} - \Phi(C)\|_2 \leq \epsilon_2$. With $m \sim \tilde{\mathcal{O}}(s \text{ polylog(n)})$ random measurements we can find $W^{(opt)}$ by solving the convex program (2.20) – (2.22) that satisfies

$$\|W^{(opt)} - C\|_F \leq C_1(\frac{1}{\sqrt{m}}\epsilon_2 + \sqrt{n}\epsilon_1), \tag{2.26}$$

with high probability. Here, C_1 is a constant. Our notation $\tilde{O}$ hides logrithmic factors in s and $\log(n)$.

We establish the recovery guarantee by showing that the collection of measurements approximately act as an isometry on sparse signals. Therefore, when the off-diagonal part of C is sparse, there is enough information in the measurements for its recovery. We formalize this statement by defining the restricted isometry property and restating a well-established theorem from the field of compressed sensing.

Definition 1 *A matrix A satisfies the restricted isometry property (RIP) of order k if there exists a $\delta_k \in (0, 1)$ such that*

$$(1 - \delta_k)\|\mathbf{x}\|_2^2 \leq \|A\mathbf{x}\|_2^2 \leq (1 + \delta_k)\|\mathbf{x}\|_2^2 \tag{2.27}$$

hold for all $\mathbf{x} \in \Sigma_k$, where $\Sigma_k = \{\mathbf{x} : \|\mathbf{x}\|_0 \leq k\}$, is the set of all k-sparse vectors[29].

Theorem 2 *(Theorem 1.2 of Ref. [30]) Suppose A satisfies the RIP of order $2k$*

with $\delta_{2k} < \sqrt{2} - 1$. *Let* $\mathbf{y} = A\mathbf{x} + \mathbf{e}$ *where* $\|\mathbf{e}\|_2 \leq \epsilon$. *Then the convex program*

$$\min \|\mathbf{x}\|_1 \text{ subject to } \|\mathbf{y} - A\mathbf{x}\|_2 \leq \epsilon \tag{2.28}$$

assumes a solution $\hat{\mathbf{x}}$ *which obeys*

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_2 \leq C_0 \sigma_k(\mathbf{x})_1 + C_1 \epsilon, \tag{2.29}$$

where $\sigma_k(\mathbf{x})_1 = \min_{\mathbf{x}' \in \Sigma_k} \|\mathbf{x}' - \mathbf{x}\|_1$, *that is the* ℓ_1 *error of the best* k*-sparse approximation of* $\mathbf{x}$, *and* C_0 *and* C_1 *are constants that only depend on* δ_{2k}.

We now show that the measurements defined in Sec. 2.2.3 satisfy RIP when the diagonals of C are fixed. It is easier to analyze the measurements if we enforce the symmetries of C by construction. Because C is real and symmetric, and its diagonals are fixed, we only need to recover the upper diagonal part of it. Note that the decay rate can be simply rewritten as $\Gamma = 2\sum_{ij} r_i C_{ij} r_j = 2\sum_i r_i^2 C_{ii} + 4\sum_{i<j} r_i r_j C_{ij}$.

Let $\text{uvec}(X)$ to be the operation that gives vectorized upper triangular part above the main diagonal of a matrix X, that is $\mathbf{x} = \text{uvec}(X)$. Therefore, Eq. (2.22) takes a simpler form

$$\|Q\mathbf{w} - \mathbf{y}\| \leq \epsilon_2 + \sqrt{mn}\epsilon_1, \tag{2.30}$$

where $\mathbf{y} = \frac{1}{4}\mathbf{h} - \frac{1}{2}\Phi(\frac{1}{2}\text{diag}(\mathbf{g}))$, $\mathbf{w} = \text{uvec}(W)$, and $Q = (q_{ij})$ is a matrix whose rows $\mathbf{Q}_i$ are given by $\mathbf{Q}_i = \frac{1}{2\tilde{p}}\text{uvec}(\mathbf{r}^{(i)}(\mathbf{r}^{(i)})^\intercal)$. Note that the elements of Q are bounded,

i.e., $|q_{ij}| \leq \frac{1}{2\tilde{p}}$. Moreover, $\mathbf{Q}_i$ are isotropic random vectors, that is

$$\mathbb{E}[\mathbf{Q}] = 0, \tag{2.31}$$

$$\mathbb{E}[\mathbf{Q}\mathbf{Q}^\intercal] = I, \tag{2.32}$$

which follows from

$$\begin{cases} \mathbb{E}[r_i r_j] = 0 & i \neq j \\ \mathbb{E}[r_i r_j r_k r_l] = 0 & i \neq j \text{ and } k \neq l \text{ and } \{i,j\} \cap \{k,l\} = \emptyset \\ \mathbb{E}[r_i r_j r_k r_l] = 0 & i \neq j \text{ and } k \neq l \text{ and } |\{i,j\} \cap \{k,l\}| = 1 \\ \mathbb{E}[r_i r_j r_k r_l] = 4p^2 & i \neq j \text{ and } k \neq l \text{ and } i = k \text{ and } j = l \end{cases}. \tag{2.33}$$

Note that in the last line of Eq. 2.33, we cannot have a case with $i = l$ and $j = k$, as the requirements of $i < j$ and $k < l$ lead to a contradiction. Therefore, using the following theorem we establish that Q satisfies the RIP.

Theorem 3 *(Theorem 5.71 of Ref. [31]) Let $A = (a_{ij})$ be an $m \times n$ matrix whose rows A_i are independent isotropic random vectors in $\mathbb{R}^n$. Let K be a number such that all entries $|a_{ij}| \leq K$ almost surely. Then the normalized matrix $\bar{A} = \frac{1}{\sqrt{m}} A$ satisfies the following for $m \leq n$, for every sparsity level $1 < k \leq n$ and every number $\delta \in (0,1)$:*

$$if\ m \geq C\delta^{-2} k \log n \log^2(k) \log(\delta^{-2} k \log n \log^2 k) \quad then\ \mathbb{E}\delta_k(\bar{A}) \leq \delta. \tag{2.34}$$

Here $C = C_K > 0$ may depend only on K.

Remark 4 *While Theorem 3 only guarantees the RIP in the expectation value, it is well known that one can obtain a version of this theorem with high probability $1 - n^{-c \log^3 k}$ using concentration of measure techniques [31, 32]. Specifically, one can obtain this bound using Theorem 6.17 in Ledoux and Talagrand [33] and Eq. (6.19) in Ref. [34] as shown by one of the authors in Ref. [35].*

Therefore, we can reach our main result from Theorems 2 and 3 and Remark 4 with our measurements Q as the matrix A in the statement of the Theorems. The RIP order k in Definition 1 should be chosen such that $2k > s$, which is achieved in expectation value by having $m \sim \tilde{O}(s \log(n))$ through Theorem 3. Then, Theorem 2 and Remark 4 guarantees the recovery of the off-diagonal part of C with high probability –at the expense of extra factors of $\log(n)$– and provide an upper-bound on the recovery error. Note that with this chosen value of m the first error term in the right hand side of Eq. (2.29) vanishes.

2.2.5 Complex C

Here we show that if C is complex the same measurements can be used to recover both the imaginary part and the real part of C.

2.2.5.1 Derivation

Let $C = V + iT$, where we is a real symmetric matrix and T is a real skew-symmetric matrix, encoding the real and imaginary part of C. Moreover, let $R =$

(R_{ij}) denote the coefficients of the Lamb shift Hamiltonian $H = \sum_{ij} R_{ij} Z_i Z_j$ that arises from interactions with a dephasing environment. We can see that $\mathcal{L}(|\mathbf{a}\rangle\langle\mathbf{b}|) = (-\Gamma_{\mathbf{ab}} + i\Omega_{\mathbf{ab}})|\mathbf{a}\rangle\langle\mathbf{b}|$, where

$$\Gamma_{\mathbf{ab}} = \frac{1}{2}(\boldsymbol{\alpha} - \boldsymbol{\beta})^\intercal V(\boldsymbol{\alpha} - \boldsymbol{\beta}), \tag{2.35}$$

$$\Omega_{\mathbf{ab}} = \frac{1}{2}(\boldsymbol{\alpha}^\intercal T\boldsymbol{\beta} - \boldsymbol{\beta}^\intercal T\boldsymbol{\alpha}) - (\boldsymbol{\alpha}^\intercal R\boldsymbol{\alpha} - \boldsymbol{\beta}^\intercal R\boldsymbol{\beta}). \tag{2.36}$$

with $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ defined as in Section 2.2.1. Therefore, the signal has an oscillatory part and a decaying part. Note that it is possible to measure $\Gamma_{\mathbf{ab}}$ and $\Omega_{\mathbf{ab}}$ by sampling the signal $\exp[(i\Omega_{\mathbf{ab}} - \Gamma_{\mathbf{ab}})]t$ different times t. Similar to spectral lines commonly encountered in atomic physics, the squared magnitude of the Fourier transform of the measurement time series is a Lorentzian function whose maximum's location gives the oscillation frequency and its half width at half maximum gives the decay rate. From an estimate of the decay rate we can recover V, the real part of C as before. We now show that the measurements of the frequencies satisfy RIP and boundedness properties. Therefore, from these measurements we can extract T, the imaginary part of C, together with the Lamb shift Hamiltonian R.

2.2.5.2 Properties of the measurements

The measurements acts on the imaginary part of C denoted by T and the Lamb shift R in the Hamiltonian as

$$\Omega_{\mathbf{ab}} = \begin{bmatrix} \boldsymbol{\alpha}^\intercal & \boldsymbol{\beta}^\intercal \end{bmatrix} \begin{bmatrix} -R & \frac{1}{2}T \\ \frac{1}{2}T^\intercal & R \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix}. \tag{2.37}$$

Again it is easier to analyze the measurements if we enforce the symmetries of R and T. Note that T and R are real skew-symmetric and symmetric matrices, respectively. Moreveover, they are both traceless. Therefore we have

$$\Omega_{\mathbf{ab}} = 2\sum_{i<j}(\alpha_i\beta_j - \beta_i\alpha_j)T_{ij} + 2\sum_{i<j}(\alpha_i\alpha_j - \beta_i\beta_j)R_{ij} \tag{2.38}$$

We can define the measurement operator $\mathbf{V}$ as a matrix whose rows $\mathbf{V}_i$ are given by

$$\mathbf{V}_i = \begin{bmatrix} \text{uvec}(\boldsymbol{\alpha}\boldsymbol{\beta}^\intercal - \boldsymbol{\beta}\boldsymbol{\alpha}^\intercal) \\ \text{uvec}(\boldsymbol{\alpha}\boldsymbol{\alpha}^\intercal - \boldsymbol{\beta}\boldsymbol{\beta}^\intercal) \end{bmatrix}. \tag{2.39}$$

Therefore, to show that $\mathbf{V}_i$'s are isotropic random vectors, we have to show that

$$\mathbb{E}[\mathbf{V}] = 0, \tag{2.40}$$

$$\mathbb{E}[\mathbf{V}\mathbf{V}^\intercal] = I. \tag{2.41}$$

Let us assume that α and β are chosen independentry at random as

$$\alpha_i, \beta_i = \begin{cases} -1 & \text{with probability } p \\ \\ 1 & \text{with probability } 1-p \end{cases} . \tag{2.42}$$

We then have

$$\mathbb{E}[\alpha_i \alpha_j] = \mathbb{E}[\alpha_i \beta_j] = \mathbb{E}[\beta_i \beta_j] = (1-2p)^2, \tag{2.43}$$

which shows that $\mathbb{E}[\mathbf{V}] = 0$.

For the covariance matrix, $\mathbb{E}[\mathbf{V}\mathbf{V}^\intercal]$, we have to consider correlations in uvec($\boldsymbol{\alpha}\boldsymbol{\beta}^\intercal - \boldsymbol{\beta}\boldsymbol{\alpha}^\intercal$), uvec($\boldsymbol{\alpha}\boldsymbol{\alpha}^\intercal - \boldsymbol{\beta}\boldsymbol{\beta}^\intercal$) and their cross correlations.

We first consider

$$\mathbb{E}[(\alpha_i\beta_j - \beta_i\alpha_j)(\alpha_k\beta_l - \beta_k\alpha_l)] = \mathbb{E}[\alpha_i\beta_j\alpha_k\beta_l - \alpha_i\beta_j\beta_k\alpha_l - \beta_i\alpha_j\alpha_k\beta_l + \beta_i\alpha_j\beta_k\alpha_l] \tag{2.44}$$

Similar to the analysis of $\mathbf{Q}$ we consider different cases. Note that in the following we assume $i < j$ and $k < l$ as we are only considering off-diagonal uppertriangular elements of V.

If $\{i, j\} \cap \{k, l\} = \emptyset$ we have

$$\mathbb{E}[\alpha_i\beta_j\alpha_k\beta_l - \alpha_i\beta_j\beta_k\alpha_l - \beta_i\alpha_j\alpha_k\beta_l + \beta_i\alpha_j\beta_k\alpha_l] = 0, \tag{2.45}$$

because of the symmetry of the terms.

If $|\{i,j\} \cap \{k,l\}| = 1$, without the loss of generality we assume that $i = k$ and find

$$\begin{cases} \mathbb{E}[\alpha_i\beta_j\alpha_k\beta_l] &= \mathbb{E}[\alpha_i\alpha_i\alpha_j\alpha_l] = \mathbb{E}[\alpha_i^2\beta_j\beta_l] = \mathbb{E}[\beta_j\beta_l] = (1-2p)^2 \\ \mathbb{E}[\alpha_i\beta_j\beta_k\alpha_l] &= \mathbb{E}[\alpha_i\beta_i\beta_j\beta_l] = (1-2p)^4 \\ \mathbb{E}[\beta_i\alpha_j\alpha_k\beta_l] &= \mathbb{E}[\beta_i\alpha_i\alpha_j\beta_l] = (1-2p)^4 \\ \mathbb{E}[\beta_i\alpha_j\beta_k\alpha_l] &= \mathbb{E}[\beta_i\beta_i\alpha_j\alpha_l] = \mathbb{E}[\beta_i^2\alpha_j\alpha_l] = \mathbb{E}[\alpha_j\alpha_l] = (1-2p)^2 \end{cases} . \tag{2.46}$$

Therefore, we find

$$\mathbb{E}[\alpha_i\beta_j\alpha_k\beta_l - \alpha_i\beta_j\beta_k\alpha_l - \beta_i\alpha_j\alpha_k\beta_l + \beta_i\alpha_j\beta_k\alpha_l] = 2(1-2p)^2 - 2(1-2p)^4. \tag{2.47}$$

Note that if we had assumed $j = l$ we would have obtained the same results.

Finally, when $|\{i,j\} \cap \{k,l\}| = 2$, we have that $i = k$ and $j = l$ and find

$$\mathbb{E}[\alpha_i\beta_j\alpha_k\beta_l - \alpha_i\beta_j\beta_k\alpha_l - \beta_i\alpha_j\alpha_k\beta_l + \beta_i\alpha_j\beta_k\alpha_l] = \mathbb{E}[\alpha_i^2\beta_j^2 - \alpha_i\beta_i\alpha_j\beta_j - \alpha_i\beta_i\alpha_j\beta_j + \beta_i^2\alpha_j^2] \tag{2.48}$$

$$= 2 - 2(1-2p)^4 \tag{2.49}$$

For the contributions from the R matrix, the elements of the measurement vector are $\alpha_i\alpha_j - \beta_i\beta_j$. Therefore, we have to consider

$$\mathbb{E}[(\alpha_i\alpha_j - \beta_i\beta_j)(\alpha_k\alpha_l - \beta_k\beta_l)] = \mathbb{E}[\alpha_i\alpha_j\alpha_k\alpha_l - \alpha_i\alpha_j\beta_k\beta_l - \beta_i\beta_j\alpha_k\alpha_l + \beta_i\beta_j\beta_k\beta_l], \tag{2.50}$$

where $i < j$ and $k < l$. Similarly, we consider different cases:

If $\{i, j\} \cap \{k, l\} = \emptyset$ we have

$$\mathbb{E}[\alpha_i \alpha_j \alpha_k \alpha_l - \alpha_i \alpha_j \beta_k \beta_l - \beta_i \beta_j \alpha_k \alpha_l + \beta_i \beta_j \beta_k \beta_l] = 0, \tag{2.51}$$

because of the symmetry of the terms.

If $|\{i, j\} \cap \{k, l\}| = 1$, without the loss of generality we assume that $i = k$ and find

$$\begin{cases} \mathbb{E}[\alpha_i \alpha_j \alpha_k \alpha_l] & = \mathbb{E}[\alpha_i \alpha_i \alpha_j \alpha_l] = \mathbb{E}[\alpha_i^2 \alpha_j \alpha_l] = \mathbb{E}[\alpha_j \alpha_l] = (1-2p)^2 \\ \mathbb{E}[\alpha_i \alpha_j \beta_k \beta_l] & = \mathbb{E}[\alpha_i \beta_i \alpha_j \beta_l] = (1-2p)^4 \\ \mathbb{E}[\beta_i \beta_j \alpha_k \alpha_l] & = \mathbb{E}[\beta_i \alpha_i \beta_j \alpha_l] = (1-2p)^4 \\ \mathbb{E}[\beta_i \beta_j \beta_k \beta_l] & = \mathbb{E}[\beta_i \beta_i \beta_j \beta_l] = \mathbb{E}[\beta_i^2 \beta_j \beta_l] = \mathbb{E}[\beta_j \beta_l] = (1-2p)^2 \end{cases}. \tag{2.52}$$

Therefore, we find

$$\mathbb{E}[\alpha_i \alpha_j \alpha_k \alpha_l - \alpha_i \alpha_j \beta_k \beta_l - \beta_i \beta_j \alpha_k \alpha_l + \beta_i \beta_j \beta_k \beta_l] = 2(1-2p)^2 - 2(1-2p)^4. \tag{2.53}$$

Finally, when $|\{i, j\} \cap \{k, l\}| = 2$, we have $i = k$ and $j = l$ and find

$$\mathbb{E}[\alpha_i \alpha_j \alpha_k \alpha_l - \alpha_i \alpha_j \beta_k \beta_l - \beta_i \beta_j \alpha_k \alpha_l + \beta_i \beta_j \beta_k \beta_l] = \mathbb{E}[\alpha_i^2 \alpha_j^2 - \alpha_i \beta_i \alpha_j \beta_j - \alpha_i \beta_i \alpha_j \beta_j + \beta_i^2 \beta_j^2] \tag{2.54}$$

$$= 2 - 2(1-2p)^4 \tag{2.55}$$

By choosing $p = \frac{1}{2}$ we can satisfy conditions (2.40) and (2.41).

We now consider cross correlations given by

$$\mathbb{E}[(\alpha_i\beta_j - \beta_i\alpha_j)(\alpha_k\alpha_l - \beta_k\beta_l)] = \mathbb{E}[\alpha_i\beta_j\alpha_k\alpha_l - \alpha_i\beta_j\beta_k\beta_l - \beta_i\alpha_j\alpha_k\alpha_l + \beta_i\alpha_j\beta_k\beta_l], \quad (2.56)$$

with $i < j$ and $k < l$. If $\{i, j\} \cap \{k, l\} = \emptyset$ we have

$$\mathbb{E}[\alpha_i\beta_j\alpha_k\alpha_l - \alpha_i\beta_j\beta_k\beta_l - \beta_i\alpha_j\alpha_k\alpha_l + \beta_i\alpha_j\beta_k\beta_l] = 0, \quad (2.57)$$

because of the symmetry of the terms.

If $|\{i, j\} \cap \{k, l\}| = 1$, without the loss of generality we assume that $i = k$ and find

$$\begin{cases} \mathbb{E}[\alpha_i\beta_j\alpha_k\alpha_l] &= \mathbb{E}[\alpha_i\beta_j\alpha_i\alpha_l] = \mathbb{E}[\alpha_i^2\beta_j\alpha_l] = \mathbb{E}[\beta_j\alpha_l] = (1-2p)^2 \\ \mathbb{E}[\alpha_i\beta_j\beta_k\beta_l] &= \mathbb{E}[\alpha_i\beta_j\beta_i\beta_l] = (1-2p)^4 \\ \mathbb{E}[\beta_i\alpha_j\alpha_k\alpha_l] &= \mathbb{E}[\beta_i\alpha_j\alpha_i\alpha_l] = (1-2p)^4 \\ \mathbb{E}[\beta_i\alpha_j\beta_k\beta_l] &= \mathbb{E}[\beta_i\alpha_j\beta_i\beta_l] = \mathbb{E}[\beta_i^2\alpha_j\beta_l] = \mathbb{E}[\alpha_j\beta_l] = (1-2p)^2 \end{cases}. \quad (2.58)$$

Therefore, we find

$$\mathbb{E}[\alpha_i\beta_j\alpha_k\alpha_l - \alpha_i\beta_j\beta_k\beta_l - \beta_i\alpha_j\alpha_k\alpha_l + \beta_i\alpha_j\beta_k\beta_l] = 2(1-2p)^2 - 2(1-2p)^4. \quad (2.59)$$

Finally, when $|\{i,j\} \cap \{k,l\}| = 2$, we have $i = k$ and $j = l$ and find

$$\mathbb{E}[\alpha_i\beta_j\alpha_k\alpha_l - \alpha_i\beta_j\beta_k\beta_l - \beta_i\alpha_j\alpha_k\alpha_l + \beta_i\alpha_j\beta_k\beta_l] = \mathbb{E}[\alpha_i^2\beta_j\alpha_j - \alpha_i\beta_i\beta_j^2 - \beta_i\alpha_i\alpha_j^2 + \beta_i^2\alpha_j\beta_j] = 0 \tag{2.60}$$

By choosing $p = \frac{1}{2}$ we can satisfy conditions (2.40) and (2.41).

2.3 Numerical examples

In this section we show how theorems developed in Sec. 2.2.3 applies in practice.

Assume that we have n qubits, whose individual dephasing rates is fully characterized. However, correlations in noise cannot be observed in single qubit measurements. We assume that there are s' pairs of qubits that are correlated. To generate the correlation matrix C that is positive semidefinite with a controllable number of non-zero offf-diagonal elements $s = 2s'$, we choose

$$\begin{cases} c_{ii} = 2 \quad i = 1, \ldots, n \\ c_{ij} = \frac{1}{2} \quad j = i \pm 1 \text{ and } 1 \leq i \leq s' \end{cases} \tag{2.61}$$

We then remove the spatial structure in the matrix by randomly permuting the rows and columns of C, that is we map c_{ij} to $c_{\pi(i),\pi(j)}$ with $\pi \in S_n$. This procedure ensures that the eigenvalues of C are non-negative. We then generate noiseless measurements by choosing m samples of $\mathbf{a}$, and $\mathbf{b}$. Finally, we assume that the diagonal elements of C are known and we solve the convex program (2.20)-(2.23)

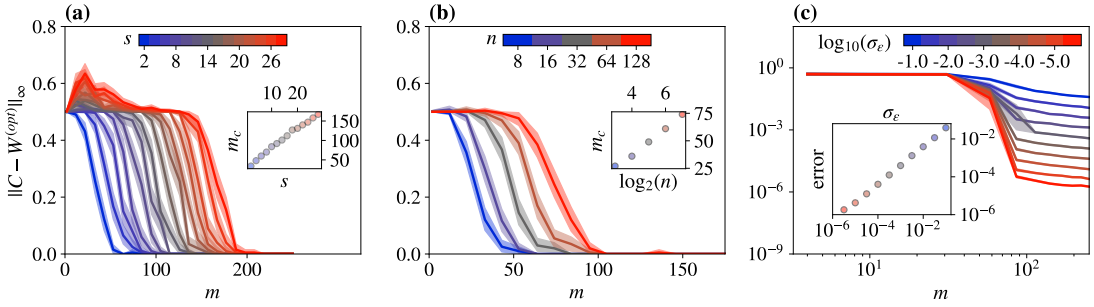

Figure 2.2: Scaling of the reconstruction error $\|W^{(opt)} - C\|_\infty$ under various circumstances. The solid lines are the average errors over 100 random instances of the problem and the shaded region is their standard deviation. (a) The reconstruction error as a function of the number of noiseless measurements m for sparsity $s = 16$ and various values of sparsity s. The errors go through a phase transitions whose location scales linearly with s. (b) The reconstruction error as a function of the number of noiseless measurements m for $n = 64$ qubits and various number of qubits n. The phase transition point scales logarithmically with n. (c) The reconstruction error as a function the number of measurements m for different values of added noise strength σ_ϵ for $(n, s) = (64, 12)$. The inset shows the errors after the transition point scale linearly with σ_ϵ as expected.

with $\epsilon_1 = \epsilon_2 = 0$ using CVXPY a convex optimization package for Python [36, 37]. In Fig. 2.2(a) we show the recovery error as a function of m for a fixed n and various choices of s. The sharp transition in the recovery error as a function of m is evident. Moreover, as shown in the inset of Fig. 2.2(a), the transition point m_c, that we define as the point where $\|C - W^{(opt)}\|_\infty$, scales linearly with s, consistent with our analytical results. In Fig. 2.2(b) we fix s, vary n, and study the recovery errors as a function of m. Again, we observe a transition in the errors as a function of m. In this case, m_c scales polynomially with $\log(n)$ as suggested in the inset of Fig. 2.2(b).

We also investigate the effect of noise on the measurements. We add a random vector $\mathbf{e}$ whose entries are independent Gaussian random variables with mean 0 and standard deviation σ_ϵ to measurement results $\frac{1}{2}\mathbf{h}$. Additionally, we assume that the diagonals of C are known and hence $\epsilon_1 = 0$. We choose $\epsilon_2 = \mathbb{E}\|\mathbf{e}\|_2 = \sqrt{m}\sigma_\epsilon$

and solve (2.20)-(2.23). The scaling of the reconstruction error $\|W^{(opt)} - C\|_\infty$ as a function of σ_ϵ is shown in Fig. 2.2 (c). The recovery error after the phase transition point scales linearly with σ_ϵ, consistent with our analytical bounds.

2.3.1 Effect of gate errors

We investigate the effect of gate errors on infering the decay rates in the experiment. Let ρ_0 and E_0 denote the noiseless initial state and observable of interest, respectively. We have

$$\rho_0 = E_0 = \frac{1}{2}(|\mathbf{a}\rangle\langle\mathbf{a}| + |\mathbf{b}\rangle\langle\mathbf{b}| + |\mathbf{b}\rangle\langle\mathbf{a}| + |\mathbf{a}\rangle\langle\mathbf{b}|). \tag{2.62}$$

We consider error channels $\mathcal{E}_s$ and $\mathcal{E}_m$ that act on state preparation and meausrement operations as

$$\tilde{\rho} = \mathcal{E}_s(\rho_0) = \rho_0 + \delta\rho \tag{2.63}$$

$$\tilde{E} = \mathcal{E}_m(E_0) = E_0 + \delta E, \tag{2.64}$$

where $\|\delta\rho\|_1 \leq \epsilon_s$ and $\|\delta E\| \leq \epsilon_m$, and $\epsilon_{s,m}$ are small parameters. The outcome of the protocol is now given by

$$\tilde{P}_+(t) = \mathrm{Tr}[\tilde{E}\mathcal{E}_t(\tilde{\rho})]. \tag{2.65}$$

where $\mathcal{E}_t = \exp(\mathcal{L}t)$ is the evolution under the correlated dephasing noise (2.1). We show that our protocol is robust against gate errors and in short times the decay of P_+ is still dominated by $\Gamma_{\mathbf{ab}}$. Using Eqs. (2.63) and (2.64) we find that

$$\begin{aligned}\mathrm{Tr}[\tilde{E}\mathcal{E}_t(\tilde{\rho})] = \mathrm{Tr}[E_0\mathcal{E}_t(\rho_0)] \\ + \mathrm{Tr}[E_0\mathcal{E}_t(\delta\rho)] + \mathrm{Tr}[\delta E\mathcal{E}_t(\rho_0)] \\ + \mathrm{Tr}[\delta E\mathcal{E}_t(\delta\rho)]\end{aligned} \tag{2.66}$$

The first term is the outcome without gate errors and we have

$$\mathrm{Tr}[E_0\mathcal{E}_t(\rho_0)] = \frac{1+e^{-t\Gamma_{\mathbf{ab}}}}{2}. \tag{2.67}$$

We can find the effect of errors on the second and third term by considering the effect of $\mathcal{E}_t$ on ρ_1 and E_1. Specifically, we find

$$\mathrm{Tr}[E_0\mathcal{E}_t(\delta\rho)] = \mathrm{Tr}[\mathcal{E}_t^\dagger(|\mathbf{a}\rangle\langle\mathbf{a}|)\delta\rho] + \mathrm{Tr}[\mathcal{E}_t^\dagger(|\mathbf{b}\rangle\langle\mathbf{b}|)\delta\rho] + \mathrm{Tr}[\mathcal{E}_t^\dagger(|\mathbf{a}\rangle\langle\mathbf{b}|)\delta\rho] + \mathrm{Tr}[\mathcal{E}_t^\dagger(|\mathbf{b}\rangle\langle\mathbf{a}|)\delta\rho] \tag{2.68}$$

$$= \mathrm{Tr}[|\mathbf{a}\rangle\langle\mathbf{a}|\delta\rho] + \mathrm{Tr}[|\mathbf{b}\rangle\langle\mathbf{b}|\delta\rho] + e^{-\Gamma_{\mathbf{ab}}t}\mathrm{Tr}[|\mathbf{a}\rangle\langle\mathbf{b}|\delta\rho] + e^{-\Gamma_{\mathbf{ab}}t}\mathrm{Tr}[|\mathbf{b}\rangle\langle\mathbf{a}|\delta\rho] \tag{2.69}$$

$$= \eta_s + \zeta_s e^{-\Gamma_{\mathbf{ab}}t} \tag{2.70}$$

where $\mathcal{E}_t^\dagger = \exp(\mathcal{L}^\dagger t)$ is the adjoint dephasing map, which coincides with $\mathcal{E}_t$ in our case as the Lindblad operators are all Hermitian. Similarly we find

$$\mathrm{Tr}[\delta E \mathcal{E}_t(\rho_0)] = \mathrm{Tr}[\delta E \mathcal{E}_t(|\mathbf{a}\rangle\langle\mathbf{a}|)] + \mathrm{Tr}[\delta E \mathcal{E}_t(|\mathbf{b}\rangle\langle\mathbf{b}|)] + \mathrm{Tr}[\delta E \mathcal{E}_t(|\mathbf{a}\rangle\langle\mathbf{b}|)] + \mathrm{Tr}[\delta E \mathcal{E}_t(|\mathbf{b}\rangle\langle\mathbf{a}|)] \tag{2.71}$$

$$= \mathrm{Tr}[\delta E|\mathbf{a}\rangle\langle\mathbf{a}|] + \mathrm{Tr}[\delta E|\mathbf{b}\rangle\langle\mathbf{b}|] + e^{-\Gamma_{\mathbf{ab}}t}\mathrm{Tr}[\delta E|\mathbf{a}\rangle\langle\mathbf{b}|] + e^{-\Gamma_{\mathbf{ab}}t}\mathrm{Tr}[\delta E|\mathbf{b}\rangle\langle\mathbf{a}|] \tag{2.72}$$

$$= \eta_m + \zeta_m e^{-\Gamma_{\mathbf{ab}}} \tag{2.73}$$

We see that these terms decay with the same rate as the first case and do not afffect the exponential decay. However, the last term can, in principle, contain different decay rates and can cause deviation from a single exponential decay. Therefore, we

bound the rate at which $R(t) = \mathrm{Tr}[\delta E \mathcal{E}_t(\delta\rho)]$ grows:

$$|\dot{R}(t)| = |\tfrac{\partial}{\partial t}\mathrm{Tr}[\delta E \mathcal{E}_t(\delta\rho)]| \tag{2.74}$$

$$= |\mathrm{Tr}[\delta E \mathcal{L}(\delta\rho)]| \tag{2.75}$$

$$\leq \|\delta E\| \|\mathcal{L}(\delta\rho)\|_1 \tag{2.76}$$

$$\leq \|\delta E\| \| \sum_{ij} L_i \delta\rho L_j^\dagger - \frac{1}{2}\{L_j^\dagger L_i, \delta\rho\}\|_1 \tag{2.77}$$

$$\leq \|\delta E\| \sum_{ij} \|L_i \delta\rho L_j^\dagger - \frac{1}{2}\{L_j^\dagger L_i, \delta\rho\}\|_1 \tag{2.78}$$

$$\leq \|\delta E\| \sum_{ij} \|L_i\| \|L_j^\dagger\| \|\delta\rho\|_1 + \frac{1}{2}\|L_j^\dagger\| \|L_i\| \|\delta\rho\|_1 + \frac{1}{2}\|L_j^\dagger\| \|L_i\| \|\delta\rho\|_1 \tag{2.79}$$

$$\leq \|\delta E\| \|\delta\rho\|_1 (\sum_{ij} 2\|L_i\| \|L_j^\dagger\|) \tag{2.80}$$

$$\leq 2\epsilon_m \epsilon_s (n+s) \tag{2.81}$$

where we used the Hölder's inequality in deriving the third line from the second line[38]. We also used the fact that $L_i = Z_i$ in our problem and $\|Z_i\| = 1$. In deriving (2.78) from (2.77) we used the following identity

$$\|AB\|_1 \leq \|A\| \|B\|_1 \tag{2.82}$$

We can obtain this inequality by first noting that for any matrix A and B, $\sigma(AB)_i \leq \sigma_{\max}(A)\sigma_i(B)$, where σ_i are the ordered singular values[39].

Now we prove that $\sigma(AB)_i \leq \sigma_i(A)\sigma_{\max}(B)$. First note that for Hermitian

matrices A and B we have:

$$A \geq B \implies \lambda_i(A) \geq \lambda_i(B) \tag{2.83}$$

where λ_i's are ordered eigenvalues, and $A \geq B$ means that $A - B$ is positive semidefinite. We then show that $\lambda_i(AB) \leq \lambda_{\max}(A)\lambda_i(B)$ if $A \geq 0$ and $B \geq 0$, that is A and B are positive semidefinite. Note that $\lambda_i(AB) = \lambda_i(B^{\frac{1}{2}}AB^{\frac{1}{2}})$, which can be seen by

$$AB|i\rangle = \lambda_i|i\rangle \implies AB^{\frac{1}{2}}B^{\frac{1}{2}}|i\rangle = \lambda_i|i\rangle \overset{\times B^{\frac{1}{2}}}{\implies} B^{\frac{1}{2}}AB^{\frac{1}{2}}(B^{\frac{1}{2}}|i\rangle) = \lambda_i(B^{\frac{1}{2}}|i\rangle) \tag{2.84}$$

Next, note that $\lambda_{\max}(A)I - A \geq 0$. Therefore, $B^{\frac{1}{2}}(\lambda_{\max}(A)I - A)B^{\frac{1}{2}} \geq 0$. This is because $\langle x|B^{\frac{1}{2}}(\lambda_{\max}(A)I - A)B^{\frac{1}{2}}|x\rangle = \langle y|\lambda_{\max}(A)I - A|y\rangle \geq 0$ for all $|y\rangle = B^{\frac{1}{2}}|x\rangle$. We then obtain

$$B^{\frac{1}{2}}AB^{\frac{1}{2}} \leq B^{\frac{1}{2}}AB^{\frac{1}{2}} + B^{\frac{1}{2}}(\lambda_{\max}(A)I - A)B^{\frac{1}{2}} = \lambda_{\max}(A)B \tag{2.85}$$

Therefore, $B^{\frac{1}{2}}AB^{\frac{1}{2}} \leq \lambda_{\max}(A)B$ together with (2.83) implies that

$$\lambda_i(AB) = \lambda_i(B^{\frac{1}{2}}AB^{\frac{1}{2}}) \leq \lambda_{\max}(A)\lambda_i(B). \tag{2.86}$$

Now for arbitrary square matrices A and B we have

$$\sigma_i(AB) = \sqrt{\lambda_i(B^\dagger A^\dagger AB)} = \sqrt{\lambda_i(A^\dagger ABB^\dagger)}. \tag{2.87}$$

Therefore, we can use (2.86) for positive-semidefinite matrices $A^\dagger A$ and $BB^\dagger$ to obtain

$$\lambda_i(A^\dagger ABB^\dagger) \leq \lambda_{\max}(A^\dagger A)\lambda_i(BB^\dagger). \tag{2.88}$$

Since both sides of the inequality are positive, we can take their square root to obtain

$$\sigma_i(AB) \leq \sigma_{\max}(A)\sigma_i(B), \tag{2.89}$$

which implies

$$\|AB\|_1 = \sum_i \sigma_i(AB) \leq \sigma_{\max}(B) \sum_i \sigma_i(A) = \|B\|\|A\|_1, \tag{2.90}$$

Therefore, we find

$$\tilde{P}_+(t) = \mathrm{Tr}[\tilde{E}\mathcal{E}_t(\tilde{\rho})] = \frac{1}{2}(1 + \eta_s + \eta_m + (1 + \zeta_s + \zeta_m)e^{-\Gamma_{\mathbf{ab}}t}) + R(t). \tag{2.91}$$

We attribute deviations from a single exponential decay to $R(t)$. Using Eqs. (2.81) and (2.91) we can see that the decay of $\tilde{P}(t)$ is dominated by $\exp(-\Gamma_{\mathbf{ab}}t)$ for evolution times $t \lesssim 1/(2\epsilon_s\epsilon_m(n+s))^{-1}$.

We also numerically investigate the effect of gate errors on the decay rate. Specifically, we generate the error channels $\mathcal{E}_s$ and $\mathcal{E}_m$ on n qubits by adding n auxiliary qubits and applying unitary operations U_s and U_m on $2n$ qubits. We control the strength of errors by a parameter Δ such that $U = \exp(i\Delta H)$, where H is a Hermitian operator on $2n$ qubits. Then, we find $|\psi_{s,m}\rangle = U_{s,m}(\rho_0 \otimes |0\rangle\langle 0|^{\otimes n})$,

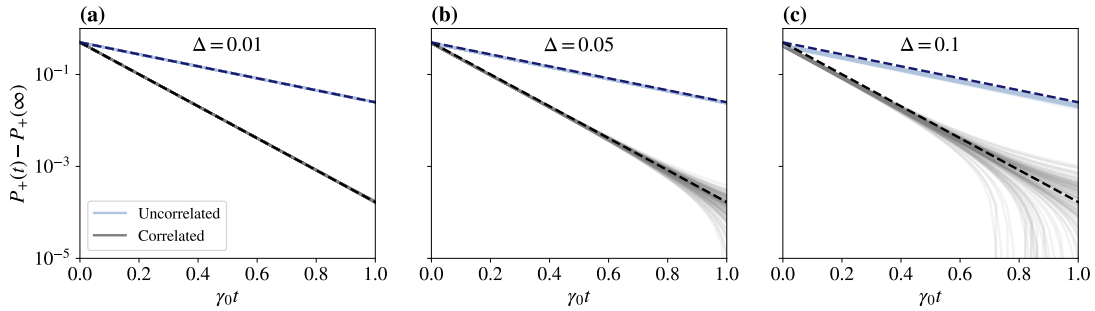

Figure 2.3: Deviations of the exponential decay for different gate errors. We simulate the decay of a 3 qubit GHZ state. We choose a correlation matrix C with $s' = 2$ in Eq. (2.61). The panels show the decay under different noise strengths: (a) $\Delta = 0.01$, (b) $\Delta = 0.05$, (c) $\Delta = 0.1$. The dashed lines show decay with no errors, and solid lines correpond to different gate error channels. The solid lines resemble the dashed lines for short evolution times. The black and blue lines indicate correlated and uncorrelated dephasing processes. These two cases are distinguishable in the presence of gate errors.

and choose $\tilde{\rho} = \mathrm{Tr}_{\mathrm{aux}}(|\psi_s\rangle\langle\psi_s|)$ and $\tilde{E} = \mathrm{Tr}_{\mathrm{aux}}(|\psi_m\rangle\langle\psi_m|)$, where the partial trace is taken over the auxiliary system. To generate different error channels we choose $H = \frac{1}{2}(M + M^\dagger)$, where elements of M are complex numbers whose real and imaginary parts are chosen uniformly at random from $[0, 1]$. We observe that the decay rate is dominated by $\Gamma_{\mathbf{ab}}$ in short times, and is distinguishable from the uncorrelated case, see Fig. 2.3.

2.4 Conclusion and outlook

We proposed a measurement protocol and a reconstruction algorithm to measure spatial correlations in the noise process that is provably efficient when the correlations are sparse. We expect that our methods can be generalized to the characterization of Markovian noise processes whose Lindblad operators are Hermitian and commuting. Moreover, our methods combined with the techniques developed in

Ref. [40] may also be used to estimate correlations in certain non-Markovian noise models more efficiently.

We expect our protocol to be useful in characterizing dephasing in the presence of weak relaxation processes, i.e. when the relaxation time is much longer than the dephasing time. However, the question of characterizing concurrent relaxation and dephasing efficiently at the generator level remains open and is a subject of future research.

While we use entangled states to probe the correlations in the noise process, it is intriguing to ask whether it is possible to achieve a similar scaling using non-entangled states. This motivates the more general question of whether entanglement can help noise spectroscopy in the same way that it is helpful in metrology[41].

Chapter 3: Improving readout fidelity of trapped-ion quantum computers with machine learning

3.1 Introduction

Quantum computing tasks involve quantum state preparation, time evolution, and measurement, accompanied by errors in all the three stages. To detect and correct errors during the evolution, quantum error correcting codes are used [42, 43, 44]. These codes rely on redundant encoding of quantum information, which makes it possible to measure syndromes and fix errors. Measurement errors not only affect the outcome of the computation, but they also limit the task of error correction. Consequently, in addition to high-fidelity operations, high quality multi-qubit readout is essential for realizing a fault tolerant quantum computer.

The quantum measurement process always involves the interaction with an external classical system. For example, collecting fluorescence from a trapped ion in a cycling transition can determine the state of the qubit [45]. In superconducting qubits, a probe signal is injected to the system through a resonator, and the phase of the output signal is used to infer the state of the qubit [46]. Spontaneous decay and excitation during the external probe can be major sources of qubit measurement

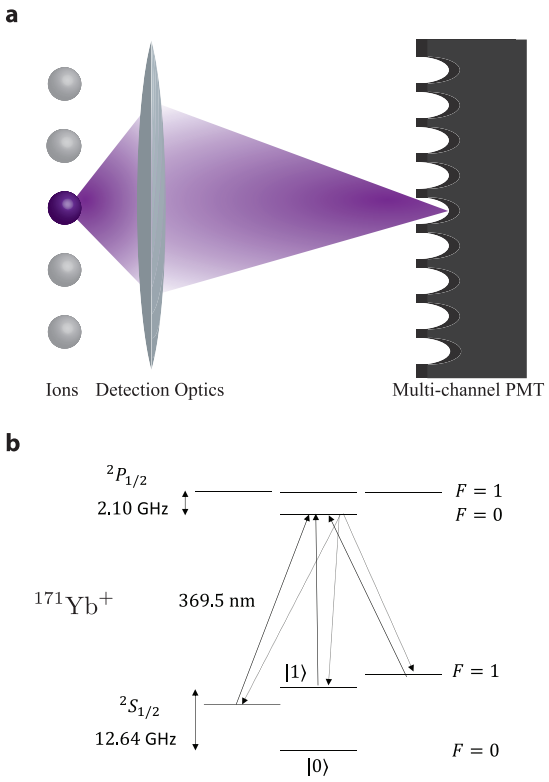

Figure 3.1: The setup and readout scheme for the trapped-ion quantum computer. (a) Schematic of our experimental setup. A single ion fluoresces inside an ion trap and its radiated photons are collected by a 0.37 NA lens. This fluorescence is then imaged onto a single channel of a multi-channel photo-multiplier tube detector. (b) Energy levels in the $^{171}Yb^+$ atomic system used for fluorescence detection. If the qubit state is $|0\rangle$, the applied detection laser is off resonance and nearly no photons are scattered. If the state is $|1\rangle$ the transition is on resonance and the ion fluoresces strongly.

errors [47]. When scaling up, the measurement signal from a qubit can be altered by the state of other qubits through crosstalk. To address this issue, one can assume an error model and infer the correct qubit states by using statistical properties of the measured data.[48].

Machine learning (ML) [49] techniques have recently become popular tools for exploring physical phenomena. For example, artificial neural networks [50] are now a powerful method for simulating the dynamics of many-body quantum systems [51,

52]. These neural networks can efficiently represent a wide class of highly correlated states [53, 54, 55], and can facilitate quantum state tomography [56]. They are also used to detect errors and decode quantum error correcting codes [57, 58, 59, 60], and to classify phases of matter [61, 62, 63]. In addition to neural networks, other ML methods, such as principal component analysis [64] and clustering [65], have been used for various tasks from classifying phases of matter [66] to discriminating measurement trajectories for improved single-qubit readout[67].

In this chapter, we exploit the versatility of ML techniques to increase the fidelity of multi-qubit measurements. While the problem of crosstalk can be partially addressed by careful statistical analysis of the data, it requires certain assumptions about the error model, which makes the integration of additional spatial and temporal features difficult. In our approach, the machine is "trained" to infer the states from the measurement results without prior knowledge of the error model. This ML method can therefore be readily generalized to other quantum computing platforms.

3.2 Initialization and Readout

We study the detection accuracy of a chain of $^{171}Yb^{+}$ ions confined in an rf Paul trap [2]. The qubit is defined by the hyperfine-split ground states of the $^2S_{1/2}$ manifold: $|0\rangle = |F = 0, m_F = 0\rangle$ and $|1\rangle = |F = 1, m_F = 0\rangle$ (see Fig. 3.1). Furthermore, we can take advantage of the $^2P_{1/2}$ level to accomplish both state preparation and measurement (SPAM) with high fidelity.

Qubit initialization is achieved by optical pumping via the $|^2P_{1/2}, F = 1\rangle$

manifold. The qubit readout, on the other hand, is performed by state-dependent fluorescence detection [47] (see Fig. 3.2). Specifically, we apply a laser beam resonant with the $|^2S_{1/2}, F = 1\rangle \rightarrow |^2P_{1/2}, F = 0\rangle$ cycling transition, and collect ion fluorescence. While the beam is on, a qubit in $|1\rangle$ will scatter photons. In contrast, a qubit in $|0\rangle$ remains dark since the light is 14.7 GHz detuned from the nearest transition with a natural linewidth of about 20 MHz. The ion fluorescence is collected by a 0.37 NA lens and each ion in the chain is imaged onto a separate channel of a 32-channel photo-multiplier tube (PMT) [2].

The histogram of the photon counts in some integration time follows a near-Poissonian distribution, centered around 0 for state $|0\rangle$ (the "dark" state) and 9 counts for state $|1\rangle$ (the "bright" state) following a 150 μs integration time. The deviations from Poissonian statistics indicate the error mechanisms in this readout scheme. The dark state histogram includes a small contribution at higher counts due to off-resonant dark-to-bright pumping during the detection step [68]. More importantly, the bright state histogram has a non-Poissonian tail towards lower photon counts due to off-resonant excitation to the $|^2P_{1/2}, F = 1\rangle$ manifold, detuned by 2.1 GHz, from which decay to $|0\rangle$ is possible [68]. By choosing an optimal collection time, 150 μs in our system, the overlap between the photon distributions corresponding to $|0\rangle$, and $|1\rangle$ can be minimized. Thus, by discriminating the two distributions one can deduce the state of the qubit. One of the commonly used techniques to distinguish between these distributions is a simple threshold discriminator, where instances with observed photon counts greater than the threshold are taken to be $|1\rangle$, and those below to be $|0\rangle$ (see Fig. 3.2). This method works very well in the single

qubit case and results in a detection fidelity, that is $\mathcal{F} = p(\text{measured x}|\text{prepared x})$, of 99.4% for $|1\rangle$, and 99.6% for $|0\rangle$, which gives an average detection fidelity of 99.5% in our setup. This error can be reduced by increasing the collection angle of the objective and reducing the detection time. A readout fidelity of $\sim 99.99\%$ has been demonstrated using this method [69].

Similar dark-to-bright or bright-to-dark pumping errors exist in other readout schemes, e.g., when a separate state with a finite lifetime is used as the dark state, known as a shelf state [48]. Smaller error contributions include laser light scattering off the ion trap and into the PMT as well as PMT dark counts, which account for 20 counts per second and 2 counts per second respectively. Both of these errors contribute only one false count for an average of 300 experiments and are therefore well discriminated using a thresholding method.

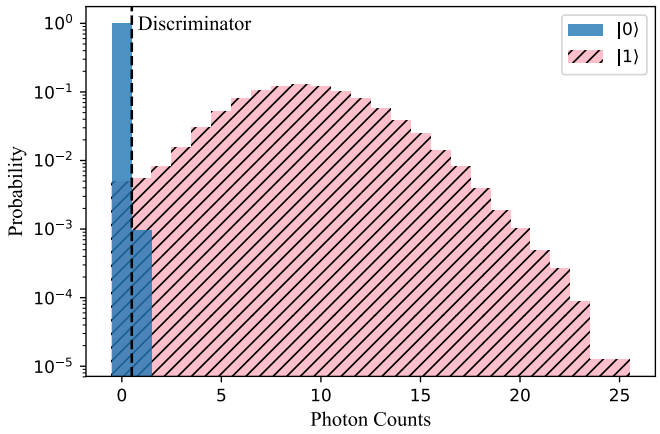

Figure 3.2: The histogram of observed photons for an integration time of 150 μs for state detection. The photon counts follow a Poisson distribution, in which the state $|0\rangle$ (solid blue) gives a mean close to zero, while the state $|1\rangle$ (shaded pink) results in nine photons on average.

When detecting the state of more than one qubit, a bright ion can cause events on other ion detector channels. This crosstalk between the PMT channels

modifies the distribution of observed photons, and the average detection fidelity decreases. One can choose a different threshold for each ion based on the state of its neighbors to partially mitigate these errors. In addition, using maximum likelihood methods, one can calculate the probability that an observed data point corresponds to the $|0\rangle$ or $|1\rangle$ state, and choose the most probable option [48, 70]. However, these methods are all tailored for a specific scenario and it is difficult to integrate other sources of information about the state, such as counts from extra PMT channels when imaging the ions onto alternating detectors, or photon arrival times. The latter contributes information about the state because bright-to-dark or dark-to-bright pumping events have characteristic photon arrival time distributions, i.e. photons arriving predominantly early or late in the detection window, which can be included in the discrimination procedure. To incorporate all data sources in a single framework and reduce the effect of crosstalk we take advantage of advances in the field of machine learning, and use an artificial neural network to perform the discrimination task. Before proceeding to the main results, we briefly introduce the neural network framework that is used in this chapter.

3.3 Methods

With N qubits, the measurement consists of photon counts and their arrival times on $M \geq N$ PMT channels. These photon counts can be binned into T time-bins to give $M \times T$ numbers that completely describe the measurements. Our goal is to classify these measurement results into 2^N states, in an N-qubit basis. Therefore,

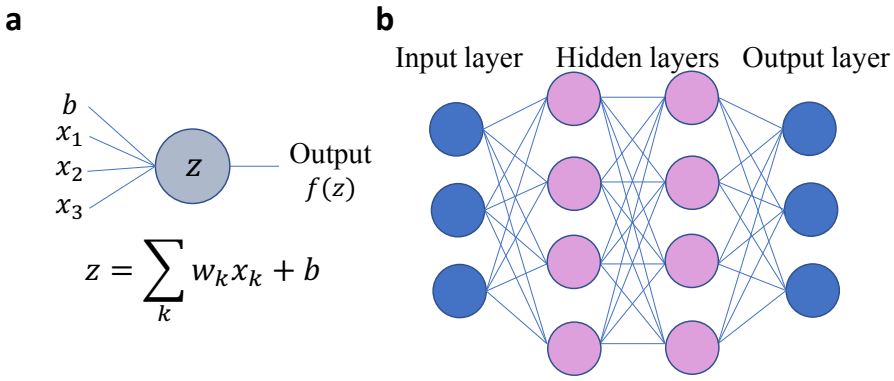

Figure 3.3: An artificial neuron and a neural network. (a) A single neuron takes inputs x_k and outputs $f(\sum_k w_k x_k + b)$, where f is called the activation function, and w_k's and b are weights and the bias of the neuron. (b) A neural network is composed of artificial neurons stacked in layers and connected to each other.

we consider a supervised learning scenario, where a set of measurement results and their corresponding states is used to train the machine and predict the correct state corresponding to a given input. We use a feed-forward neural network as depicted in Fig. 3.3 [71]. The network is built from a collection of neurons arranged in layers (columns in Fig. 3.3). A neuron is a unit that takes the input values x_k, and evalutes $f(\sum_k w_k x_k + b)$, where f is a mathematical function called the activation function, and w_k and b are scalar parameters that are referred to as neuron's weights and bias, respectively. There are various choices for the function f. A common example is the rectifier $f(z) = \max(0, z)$. When neurons are arranged in layers, the output of the previous layer serves as the input to the next layer. We represent weights and biases of the ith layer by the matrix $\mathbf{W}^{(i)}$ and the vector $\mathbf{b}^{(i)}$, respectively. In this notation, the rows of $\mathbf{W}^{(i)}$ represent the weights of the neurons in layer i, and elements of the vector $\mathbf{b}^{(i)}$ are the correponding biases. Therefore, we have $\mathbf{x}^{(i+1)} = f(\mathbf{z}^{(i)})$, where $\mathbf{z}^{(i)} = \mathbf{W}^{(i)}\mathbf{x}^{(i)} + \mathbf{b}^{(i)}$, and the function f is applied element-wise. The first

layer is called the input layer, where the neurons output the input data. Here, we have $M \times T$ neurons representing integrated photon counts from each ion in a time-bin (pixel values in Fig. 3.4). The last layer of the network is called the output layer. This layer captures different classes (states) that the input can take, that is it models $p(\text{state } i|\text{input})$. Here, we use one-hot encoding to represent the classes n, corresponding to the 2^N different quantum states in our system (image labels in Fig 3.4). That is, for each sample s, $y_{s,n} = 1$, if the sample is prepared in the nth state, and $y_{s,n} = 0$, otherwise. For classifying data into exclusive classes, it is common to use the softmax activation at the output layer, i.e. $f(\mathbf{z})_n = \frac{e^{z_n}}{\sum_k e^{z_k}}$. With this choice of activation, the output is normalized and can be interpreted as the probability of the input belonging to class n. The output with the highest probability is chosen as the predicted class. All the layers between the input and output are called hidden layers, and we use the rectifier function for them.

In order to predict the correct class, i.e., the quantum state, associated with each input, the network has to be trained. For example, with 3 ions and 7 PMT channels, the machine should predict that the input $\mathbf{x}_s^{(\text{in})} = (1\ 10\ 2\ 1\ 1\ 10\ 1)^T$ corresponds to the state $|101\rangle$, which is the 6th class and is represented by $\mathbf{y}_s = (0\ 0\ 0\ 0\ 0\ 1\ 0\ 0)^T$. The performance of the network can be quantified by a cost-function that measures the difference between the network's prediction and the target. Hence, the task of training is to find the weights and biases that optimize this cost-function. In this chapter, we use the ADADELTA optimizer [72] to minimize the cross entropy

$$C(\{\mathbf{W}^{(i)}\},\{\mathbf{b}^{(i)}\}) = -\sum_{s=1}^{N_s}\sum_{n=1}^{2^N} y_{s,n}\log(\tilde{y}_{s,n}) \tag{3.1}$$

where $\{\mathbf{W}^{(i)}\}$ and $\{\mathbf{b}^{(i)}\}$ are the sets of network weights and biases, respectively. N_s is the number of samples, 2^N is the number of classes, and $\tilde{y}_{s,n}$ is the output value of the network corresponding to the class n for the sample s.

To design and train the network, we split the data in three sets: training (60%), cross-validation (20%), and test sets(20%). The training set is used to train a given network and find its optimal weights and biases. The cross-validation set is used to evaluate the performance of networks with different number of hidden layers and neurons to choose the optimal network architecture. We observe that networks with two hidden layers perform the best. For each network, all hidden layers have the same number of neurons. This number varies from 8, for the simplest case, to 40, for the network with the most features. Such an architecture is complex enough to correctly classify the measurement data without overfitting it. Lastly, the reported performance of the optimal networks is evaluated, using the previously unseen test set.

3.4 Results

We now discuss the results in detail. We begin by moving a single trapped ion to the positions that would be occupied by ions in the multi-ion chain that we wish to investigate. This method allows us to to recreate the experimental setup with N qubits. We typically image ions onto alternating PMT channels to reduce

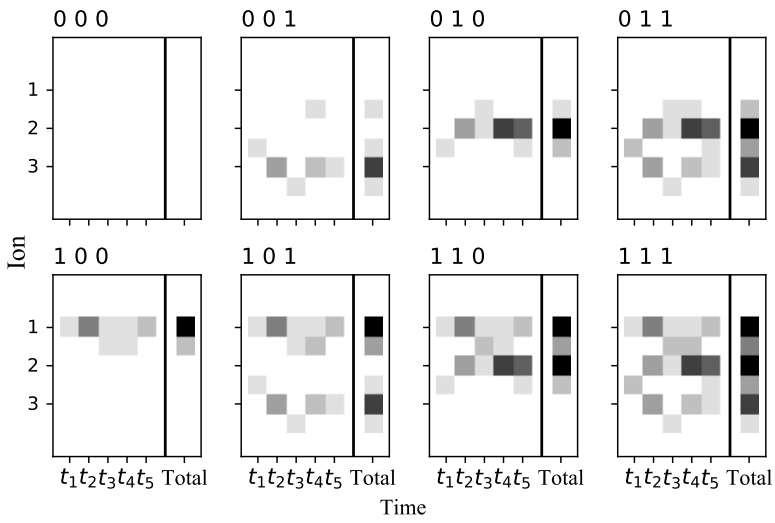

Figure 3.4: Results of an instance of the experiment for different initial states. Photon counts for all eight computation basis states over three qubits (000,001,...,111) are binned into 30 μs time-bins in a total of 150 μs collection. The last column shows the total counts and darker color indicates more photons. The effect of crosstalk is visible in the intermediate channels (unlabeled rows).

the crosstalk, which leaves the intermediate channels unused. We also take data imaging them onto neighboring PMT channels in order to explore how detection errors would change for a chain of ions with smaller inter-ion distance. Then, we either initialize the ion in $|0\rangle$ to take data on dark states or we use a high-fidelity microwave pi-pulse to create $|1\rangle$ for bright-state data. Finally, we detect the qubit state by counting how many photons are detected on the ion's corresponding PMT channel as well as neighboring channels. In addition, the photon arrival time is recorded with sub-μs resolution. By loading only a single ion, we can create the full statistics for all the 2^N computational basis states by superimposing these individual distributions. This procedure separates SPAM errors from other systematic errors present in the system such as addressing crosstalk errors. The average detection

fidelity, which includes a small error from state preparation, is given by

$$\bar{\mathcal{F}} = \frac{1}{2^N} \sum_i p(\text{measured } i | \text{prepared } i), \tag{3.2}$$

where the sum is carried over all the 2^N computational basis states.

We compare six different methods and show that machine learning approaches outperform the two commonly used strategies in state discrimination. Below we describe these six strategies:

(i) Fixed threshold (FT): A threshold for photon counts is chosen to maximize the discrimination between bright and dark probability distributions. The same threshold is used for all the ions. In experiments with more than one qubit, this threshold is higher than the single qubit case because of crosstalk. Additional background noise from superimposing the statistic of individual qubits do not significantly contribute to errors.

(ii) Adaptive threshold (AT): The threshold for each ion depends on the state of its neighbors. First, the state is determined by a fixed threshold, and then the inference process is iterated based on the state of neighbors and the corresponding thresholds.

(iii) Neural network (NN): First the photon counts from the ion PMT channels and their corresponding 2^N states (classes) are fed into a neural network. After the training, the neural network can predict the state of a given array of photon counts.

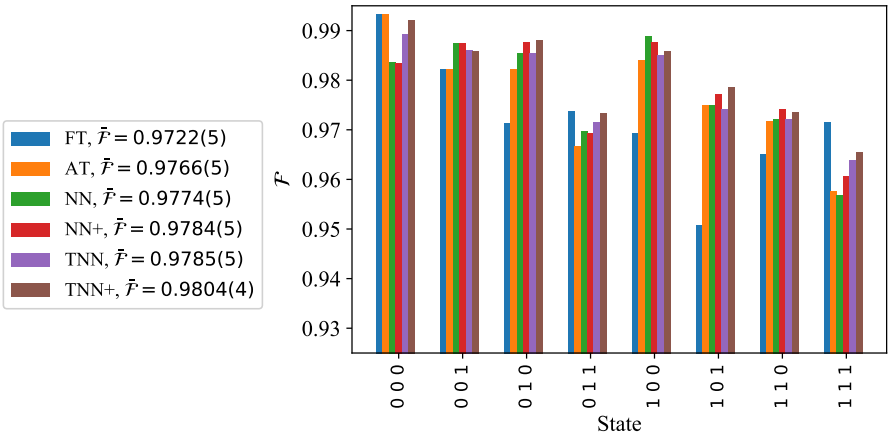

Figure 3.5: Comparison of different methods for state detection defined in the main body. We can see that the neural network (NN) methods outperform the conventional thresholding (*FT*, *AT*) methods. In addition, the performance is enhanced gradually as we provide the neural network with more features, e.g. intermediate channels (indicated by + at the end of label names) and time stamps (indicated by T at the beginning of label names). The errors given in parentheses are statistical.

(iv) Neural network with intermediate channels (*NN+*): Similar to *NN*, but the input also contains the intermediate PMT channel's data.

(v) Neural network with time-stamped data (*TNN*): The photon counts from the ion PMT channels are collected into time-bins to form a 2D image, where one axis is the time, and the other represents the location of the ions. The color intensity then represents the number of photons observed in that time-bin (see Fig. 3.4). These images with their corresponding labels are used to train the neural network.

(vi) Neural network with time-stamped data and intermediate channels (*TNN+*): The time-binned photon counts of the ion PMT channels and the intermediate channels are used to form an image, which the neural network learns to classify. This is the most comprehensive information available about the experiment.

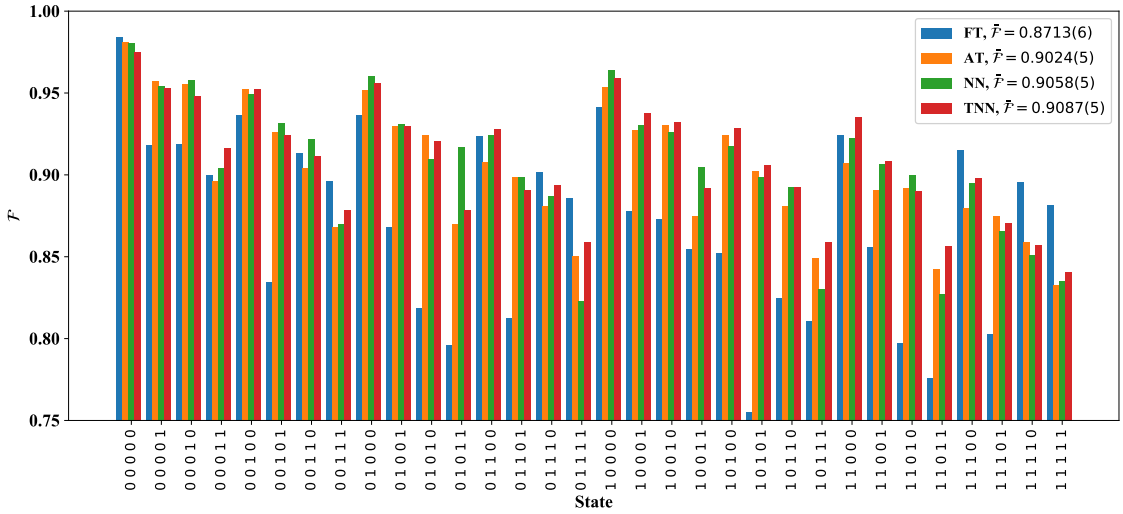

Figure 3.6: Comparing neural networks with threshold methods for five-qubit state detection. In this case the intermediate channel data is not available, but neural networks can still perform better than threshold methods. The errors given in parentheses are statistical.

We note that due to a large overlap between the photon count distributions of the intermediate channels with different bright neighbors, it is not possible to utilize the intermediate channel data with a simple threshold method. The same is true for the time-binned data, where the overlap of the bright and dark distributions is significant. This is because the distributions are Poissonian and have close mean values. However, the neural network can easily incorporate all the features and extract the available information.

In the first experiment, we consider a three qubit measurement scenario where the data from intermediate channels is available. We collect 80000 samples for each class and apply the six strategies and observe that the neural network outperforms the other methods. In Fig. 3.5, it can be seen that with the same amount of information *NN* outperforms *FT* and *AT*, and when additional information is provided *TNN+* can improve the errors over *FT* and *AT* by 30% and 17% respectively. It

can also be seen that the neural network reduces the false positives in detecting 000 and 111 states and improves the crosstalk errors in the other states. Note the architecture of the neural network is kept the same and only the number of neurons are increased to represent the more complicated features, therefore providing a flexible tool for inferring properties of the system from experimental data.

In the second experiment, the ions are moved closer to each other to represent experiments where there are many ions in the trap, and neighboring PMT channels are associated with different ions. In this case, the data from intermediate channels is no longer available, and the crosstalk errors are increased. We consider a five qubit measurement scenario with 50000 samples for each class and compare *FT*, *AT*,*NN* and *TNN* methods.

As shown in Fig. 3.6 we can see that the same behavior observed in the first experiment persists, and neural networks beat threshold methods, and incorporating time stamped data further improves the fidelities. Specifically, we observe 29% and 6% improvement by *TNN* over *FT* and *AT*, respectively.

In addition, we employ a recurrent neural network (RNN) as an alternative approach. These networks are tailored towards studying sequences of data (time-bins in our case), where the output in each step depends on the history through the internal state and an external input (see Fig. 3.7 inset). This feedback and memory effect is useful in capturing correlations in the sequence. While we observe the same fidelity as *TNN+*, this method is advantageous for experiments with variable detection time, since it can handle data sequences with different lengths. We illustrate this capability by training a long short-term memory (LSTM) network, which is a

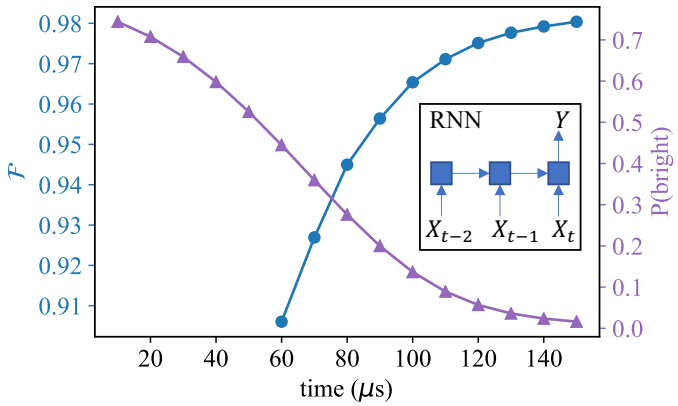

Figure 3.7: Recurrent neural network approach. The inset shows a schematic representation of the network, where the carried internal state and the output is affected by additional sequential inputs (time-binned photon counts). Left ordinate, blue circles: Performance of a recurrent network for different detection times. The fidelity increases with longer detection times. Right ordinate, purple triangles: The average probability of the ion being bright decreases with the arrival time of the first photon.

type of RNN [73], with the full sequence of measurement data, using finer time-bins of 10μs. Then, we evaluate the performance of the network by varying the length of the test sequence, and observe that indeed the performance increases with the measurement time (see circles in Fig. 3.7). In addition, we interrogate the network with artificial data to map out its internal mechanism. Specifically, we construct sequences with a single photon count, the arrival time of which is scanned. The output indicates the significance of the photon arrival time in deciding the state of the qubits. We observe that the network learns that photons with late arrival times are more likely to come from ions prepared in the dark state, which is consistent with our physical understanding of error mechanism by off-resonant excitation (see traingles in Fig. 3.7).

3.5 Conclusion and outlook

We have shown that a simple neural network classifier can improve the detection fidelities over tradition thresholding methods. The neural network classifier does not require assumptions about the system and can incorporate different data sources in one framework in a straightforward way. As the ion-trap systems are very clean and the measurements are well-described theoretically, we do not expect neural networks to beat complicated models that take into account possible errors and evaluate the likelihood of a state corresponding to measured values. Similarly, we were not able to observe significant improvements over feed-forward networks by using more sophistaced techniques such as RNNs or convolutional neural networks. This is because the patterns and correlations in the data are simple and hence well-captured by feed-forward networks. However, we expect such advanced techniques to be especially useful in other systems like superconducting qubits, where the measurement processes are more complicated and the data has intricate features.

In addition, we have considered exclusive classes in our classifier, which implies the size of the network scales exponentially with the number of qubits. However, we have observed that multi-label classifiers can achieve a performance close to our method, while maintaining a linear scaling with the number of qubits. Moreover, while it is not necessary in the current setup, by taking advantage of the locality of the crosstalk one can bootstrap smaller networks over a few qubits. By taking majority vote over the outcome of the smaller classifiers, the most probable state corresponding to the measurement results over many qubits is decided.

In conclusion, we expect that techniques such as the one presented can simplify and improve the future experiments and serve as a straightforward alternative for optimizing the readout of quantum computers as they are scaled up to many qubits.

Chapter 4: Machine learning thermodynamic arrow of time

4.1 Introduction

While the microscopic dynamics of physical systems are time reversible, the macroscopic world clearly does not share this symmetry. If we are shown a video of a macroscopic process, it is typically easy to guess whether the movie is played in the correct or in time-reversed order. In 1927, Sir Arthur Eddington coined the phrase "time's arrow" to express this asymmetry in the flow of events, arguing that it traces back to the second law of thermodynamics [74]. Nearly a century later, advances in statistical mechanics have extended our understanding of this problem to the microscopic regime. Here, fluctuations prevent us from discerning the direction of time's arrow with certainty [17, 20, 75]. Instead, the probability that a movie is being shown in the correct chronological order is determined by the energy dissipated during the process, as expressed by Eq. 4.2 below. This prediction, which has been verified experimentally [21], is equivalent to Crooks's fluctuation theorem [18, 76], an important result in modern non-equilibrium statistical mechanics [17, 19]. The recent success of applications of machine learning (ML) and artificial intelligence (AI) in physics begs the question whether these techniques can speed up scientific discovery [23, 77]. ML methods have emerged as exciting tools to study problems

in statistical and condensed matter physics, such as classifying phases of matter, detecting order parameters, and generating configurations of a system from observed data [55, 61, 63, 66, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90].

In this Chapter, we apply machine learning to the problem of time's arrow, within the framework of nonequilibrium statistical mechanics. We show that a machine can learn to accurately guess the direction of time's arrow from microscopic data and – more importantly – that it does so by effectively discovering the underlying thermodynamics, identifying dissipated work as the relevant quantity and correctly establishing its relation to time's arrow. Remarkably, the main machine learning tool used here, logistic regression (LR), was developed decades before the derivation of fluctuation theorems by human experts [91, 92]. This suggests that a data-driven approach could have led to an earlier discovery of these theorems. Moreover, we show that the machine can generate representative trajectories for forward and backward time directions correctly. Finally, we design a neural network that can detect the underlying process and classify the direction of time's arrow at the same time.

We first introduce the relevant physical laws governing microscopic, nonequilibrium fluctuations, and briefly review the ML techniques that we will use. We then apply our methods to various model physical examples and we study the ability of ML techniques to learn and quantify the direction of time's arrow.

4.2 Thermodynamics and the arrow of time

When small systems undergo thermodynamic processes, fluctuations are non-negligible and the second law is expressed in terms of averages. Thus the Clausius inequality relating the work W performed on a system to the net change in its free energy, ΔF, takes the form

$$\langle W \rangle \geq \Delta F, \tag{4.1}$$

where the angular bracket denotes an average over many repetitions of the process. These non-equilibrium fluctuations satisfy strong constraints that allow us to rewrite such inequalities in terms of stronger equalities [18, 76, 93, 94, 95], and to quantify the direction of time's arrow as a problem in statistical inference [17, 20, 76, 96, 97]. To frame this problem, let us first specify the class of processes we will study, and introduce the notation.

Consider a system in contact with a thermal reservoir at temperature β^{-1}. The system's Hamiltonian $\mathcal{H}_\lambda(\mathbf{x})$ depends on both the system's microstate $\mathbf{x}$, and a parameter λ. An external agent performs work by manipulating this parameter. Now imagine that the system begins in equilibrium with the reservoir, and then the agent varies the parameter according to a schedule $\lambda_{\mathrm{F}}(t)$ from $\lambda_F(0) = A$ to $\lambda_F(\tau) = B$. We refer to this as the *forward process*. The trajectory describing the system's evolution can be pictured as a movie, and is denoted by $\{\mathbf{x}_{A\to B}(t)\}$, where the time interval $0 \leq t \leq \tau$ is implied. We refer to this as *forward trajectory*, see Fig. 4.1(a). We also imagine the *reverse process*, in which the system starts in

an equilibrium state at $\lambda = B$, and the agent varies the parameter from B to A according to $\lambda_{\rm R}(t) = \lambda_{\rm F}(\tau - t)$. The trajectory (movie) for this process is denoted by $\{\mathbf{x}_{B\to A}(t)\}$. Finally, consider the time reversal of this trajectory, $\bar{\mathbf{x}}_{B\to A}(t) = \mathbf{x}^*_{B\to A}(\tau - t)$, where the '$*$' implies negation of momentum coordinates. This time-reversed trajectory corresponds to a movie of the reverse process, played backward in time and is referred to as *backward trajectory*, see Fig. 4.1(a).

Guessing the direction of time's arrow can be cast as a game in which a player is shown either a forward or a backward trajectory – thus in either case the player only "sees" the parameter being varied from A to B. The player must then guess which process, forward or reverse, was actually used to generate the trajectory [98]. The player's score, or accuracy, is the ratio of correct predictions to the total number of samples.

To optimize the likelihood of guessing correctly, it suffices for the player to know the sign of the quantity $W - \Delta F$, where W is the work performed on the system and $\Delta F = F_B - F_A$ is the free energy difference between its initial and final states, as depicted in the movie. Specifically, let $P(\mathrm{F}|\{\mathbf{x}(t)\})$ denote the likelihood that a given trajectory, $\{\mathbf{x}(t)\}$, is obtained by performing the forward process, and let $P(\mathrm{R}|\{\mathbf{x}(t)\})$ denote the likelihood that the trajectory is the time reversal of a realization of the reverse process. Note that $P(\mathrm{F}|\{\mathbf{x}(t)\}) + P(\mathrm{R}|\{\mathbf{x}(t)\}) = 1$. In addition, assume that the game is unbiased, e.g. the choice of performing the forward or reverse process in the first place was decided by flipping a fair coin. Then the likelihood that the trajectory was generated during the forward process is given

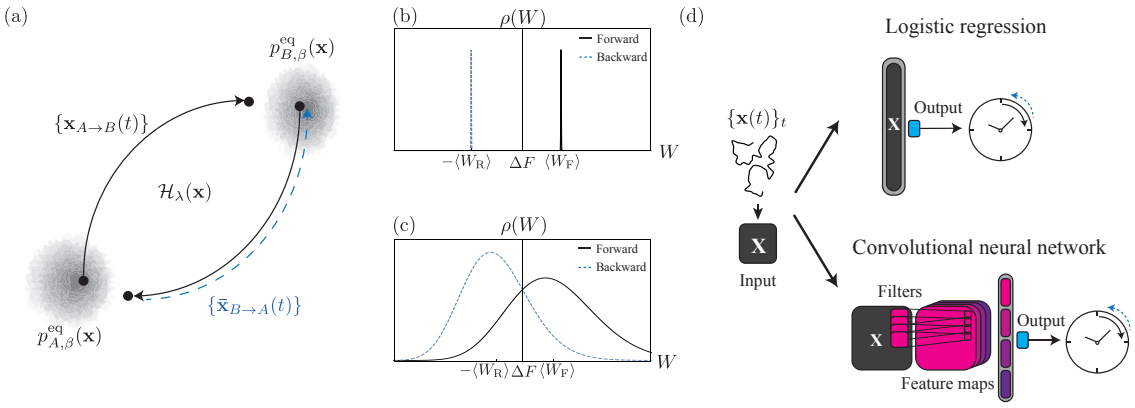

Figure 4.1: Non-equilibrium physics, time's arrow, and machine learning. (a) The system evolves under a Hamiltonian $\mathcal{H}_\lambda$ that depends on a parameter λ. The solid black trajectories depict the system's evolution during the forward ($A \to B$) and reverse process ($B \to A$). In the forward (reverse) process the system starts in equilibrium $p^{\rm eq}_{A(B),\beta}$ and λ is varied from $A(B)$ to $B(A)$, respectively. The dashed blue trajectory $\{\bar{\mathbf{x}}_{B\to A}(t)\}$ is the time-reversal of the system's evolution during the reverse process. (b) The work distribution $\rho(W)$ corresponding to the forward $W_{\rm F}$ and the backward $-W_{\rm R}$ trajectories. The change in the free energy during the forward process is denoted by ΔF. For macroscopic irreversible phenomena, fluctuations are negligible, $W_{\rm F} > \Delta F > -W_{\rm R}$, and the distinction between the forward and backward trajectories are clear. (c) Similar to (b), but for a microscopic system. Fluctuations are more pronounced compared to (b), and the distinction between the two distributions is less clear. (d) A trajectory is represented by a matrix $\mathbf{X}$. This matrix is the input to a neural network which determines the direction of time's arrow. The top shows logistic regression network, where the input is flattened and reshaped into a vector, and the output is calculated by applying a non-linear function to a linear combination of the input coordinates. The bottom shows a convolutional neural network, where at first filters are convolved with the input, making feature maps that encode abstract information about the local structure of the data. Then these feature maps are reshaped and processed through a fully-connected layer. The output of the network is used to decide the direction of time's arrow.

by [17, 96, 97]

$$P(\mathrm{F}|\{\mathbf{x}(t)\}) = \frac{1}{1+e^{-\beta(W-\Delta F)}}, \tag{4.2}$$

which is greater than (less than) 50% when $W - \Delta F$ is positive (negative). Here, the work performed by the external agent is

$$W = \int_0^\tau dt \dot{\lambda} \frac{\partial \mathcal{H}_\lambda(\mathbf{x})}{\partial \lambda}, \tag{4.3}$$

and the change in free energy is given by

$$\Delta F = -\frac{1}{\beta} \log \left(\frac{Z_{B,\beta}}{Z_{A,\beta}} \right), \tag{4.4}$$

where

$$Z_{\lambda,\beta} = \int d\mathbf{x} \exp[-\beta \mathcal{H}_\lambda(\mathbf{x})] \tag{4.5}$$

is the partition function. In macroscopic systems, the values of work performed on the system corresponding to forward trajectories, W_F, and for backward trajectories, $-W_\mathrm{R}$, are sharply peaked around their mean values, Fig. 4.1(b), and the sign of $W - \Delta F$ is a reliable indicator of the direction of time's arrow. (Here, W_R is the work performed during a given realization of the reverse process, therefore for the corresponding backward trajectory the work value is $-W_R$.) However, for microscopic systems these distributions can overlap significantly, as in Fig. 4.1(c). Eq. 4.2 shows that the player optimizes the chance of success simply by guessing "forward" whenever $W > \Delta F$, and "reverse" otherwise, without accounting for any further

details of the trajectory. Note that if $|W - \Delta F| \gg k_B T$ then determining the arrow of time is easy, but when $|W - \Delta F| \lesssim k_B T$ the problem becomes more difficult – in effect, time's arrow is blurred.

4.3 Neural networks

To train a computer program to infer the direction of time's arrow from a movie of the system's trajectory, we first simulate a number of trajectories from the forward and the reverse processes, and we "time-reverse" the latter so that each trajectory is chronologically ordered with λ varying from A to B. To generate trajectories we closely follow Ref. [76]. We consider a discrete set of time steps $t \in \{0, 1, \ldots, \tau\}$. The value of the control parameter and the state of the system at each time step is denoted by λ_t and $\mathbf{x}_t$, respectively. In the forward process, the initial state of the system is drawn from equilibrium with $\lambda = \lambda_0$. The time evolution can be broken into two substeps:

1. With the state of the system fixed, the control parameter is changed $\lambda_t \to \lambda_{t+1}$

2. At fixed λ_{t+1}, the state of the system evolves $\mathbf{x}_t \to \mathbf{x}_{t+1}$

Here, the second substep is either generated by a stochastic differential equations (for the Brownian particle) or Metropolis algorithm (for spin examples). The total work performed in this process is

$$W = \sum_{t=0}^{\tau-1} [\mathcal{H}_{\lambda_{t+1}}(\mathbf{x}_t) - \mathcal{H}_{\lambda_t}(\mathbf{x}_t)] \tag{4.6}$$

For producing backward trajectories, the system is initialized in an equilibrium state with $\lambda = \lambda_\tau$. The dynamics begin with a change in the system state, followed by a change in λ. In the end, the history of the system state is reversed, and the calculated work is negated to obtain backward trajectories and their corresponding work values. We then attach a label $y = 0$ (reverse) or $y = 1$ (forward) indicating which process was used to generate that trajectory. We then provide the machine with this collection of labelled trajectories, which serves as the training data. *A priori*, any one of the trajectories could have been generated from either the forward or the reverse process, and the training stage now consists of using a neural network (NN) classifier to construct a model of the function $P(\mathrm{F}|\{\mathbf{x}(t)\})$, which gives the likelihood the trajectory was generated by the forward process. Although this function is known analytically, Eq. 4.2, the machine is not provided with this information. We now sketch how the training is accomplished.

Since each numerically generated trajectory consists of a discretized time series of microstates, we represent the trajectory as a matrix $\mathbf{X}$ whose rows correspond to different times, and whose columns correspond to phase space coordinates. The training stage amounts to designing a function that maps any such matrix $\mathbf{X}$ onto a real number p between 0 and 1, whose value is the machine's best estimate of the likelihood that the trajectory was generated by the forward process.

In this work, we consider two types of classifiers: (i) logistic regression (LR), which can be thought of as single layer neural network, and (ii) convolutional neural network (CNN). The input to LR (the top panel of Fig. 4.1(d)) is a vectorized trajectory $\mathbf{a} = \mathrm{vec}(\mathbf{X})$, and the output is $p = g(\mathbf{\Omega}^\intercal \mathbf{a} + b)$, where $\mathbf{\Omega}$ is a vector of

weights, b is the bias, and $g(z) = 1/(1 + \exp(-z))$ is the logistic sigmoid function. The CNN can compute more complicated functions than the LR [99]. The input to our CNN is a trajectory matrix $\mathbf{X}$, and the output is again a value p. The CNN has convolutional layers that extract useful information by taking advantage of the temporal and spatial structure of the data, see the bottom panel of Fig. 4.1(d). A convolution layer convolves the input with a number of filters, then applies a non-linear function to the output of the filters. Each convolution operation with a kernel $\mathbf{\Omega}$ and bias b, maps an input matrix $\mathbf{X}$, to another matrix $\mathbf{Z} = \mathbf{\Omega} * \mathbf{X}$ given by [99]

$$Z_{j,k} = \sum_{m,n} X_{j\times s+m,k\times s+n}\Omega_{m,n} + b \tag{4.7}$$

where s specifies the number of steps the filter moves in each direction. It is called the stride of the convolution and is a hyperparameter that is tuned using the cross-validation data. The output of the convolution layer is obtained by applying a non-linear function g element-wise to $\mathbf{Z}$. The convolution layers can be repeated many times, and combined with pooling layers where the dimension of the output is reduced through a procedure such as averaging. At the end, the output of the convolution layer is flattened to form a vector and that vector is fed into a series of fully connected layers to produce the network's output [99].

The CNN's that we consider has four 2×2 filters, with the stride of 1, and with periodic boundary condition. We choose the rectifier, i.e., $g(z) = \max(0, z)$, for the activation of these filters. The output of all the filters is then combined to form a single vector. For the CNN classifying the $\mathcal{J}$-protocol (see Fig. 4.4), this

vector is fed into a single neuron with sigmoid activation, whose values determine the direction of time's arrow. For the gating network (Sec. 4.5), this vector is fed into a fully connected layer with 50 hidden neurons and the rectifier activation, followed by the output neuron with the sigmoid activation.

To train the classifier, we determine optimal values of parameters (such as the weights and biases in LR) by minimizing the cross-entropy using a gradient based optimization algorithm[99]

$$C = -\frac{1}{N_{\text{samp}}} \sum_m [y_m \log(p_m) + (1 - y_m) \log(1 - p_m)] \tag{4.8}$$

The sum is carried over the N_{samp} training samples, $y_m \in \{0, 1\}$ is the label attached to the m'th trajectory, indicating which process was used to generate the trajectory, and p_m is the output of the network for that trajectory.

Throughout this work, we always split a given data-set into three parts. We use 60%, 20% and 20% of the data for training, validation, and testing the model, respectively. We use a data-set with total of 20000 samples for the Brownian particle. For the spin chain examples ($\mathcal{B}$ and $\mathcal{J}$-protocols), we use 20000 samples for each temperature. The samples are then split into three sets and are used to train, validate, and test the models. The validation set is used to tune the architecture and hyperparameters of the model, while the test data is used for unbiased evaluation of the the final model's accuracy. We use Adam optimizer with parameters suggested in the original paper for the training [100]. We assess the performance of the network by testing it over a balanced set of trajectories, i.e. half forward and half backward.

If $p_m \geq 0.5$ then the algorithm guesses that the trajectory was generated from the forward process, otherwise it guesses the reverse process. As a figure of merit, we consider the accuracy, i.e. the ratio of correct guesses to total number of samples.

To reduce overfitting it is helpful to include a regularization term. This will help to reduce the difference between the training error and the test error. We consider L_2 regularization $\alpha \sum_\ell \Omega_\ell^2$, that is adding the square of all the weights in the network to the cost function. The parameter α is a hyper-parameter of the model and is tuned using the cross-validation data.

Model	α
Brownian particle	0.001
$\mathcal{B}$-protocol	10^{-4}
$\mathcal{J}$-protocol (LR)	2×10^{-5}
$\mathcal{J}$-protocol (CNN all layers)	10^{-4}
Coarse-grained $\mathcal{B}$ and $\mathcal{J}$-protocols	2×10^{-5}
Gating network (conv. and the hidden layer)	10^{-5}
Gating network (output)	2×10^{-5}
Alternative activation functions (all layers)	10^{-5}

Table 4.1: The value of L_2 regularization parameter α for the NNs in this Chapter. To reduce overfitting it is helpful to include a regularization term. This will help to reduce the difference between the training error and the test error. We consider L_2 regularization $\alpha \sum_\ell \Omega_\ell^2$, that is adding the square of all the weights in the network to the cost function.

Additionally, for training the CNNs in this work, we use the dropout technique to reduce overfitting. Dropout refers to deactivating and ignoring certain neurons during the training phase. Specifically, at every training step, a random fraction of p_{drop} of neurons are deactivated [101].

We find that our results do not vary significantly with the choice of hyper-parameters. We choose $p_{\text{drop}} = 0.25$ for the dropout rate of neurons of the convolu-

tional layer in the $\mathcal{J}$ network, and $p_{\text{drop}} = 0.5$ for the gating network in the mixture of experts network of Section 4.5.3. The L_2 regularization rates α are shown in Table 4.1.

4.4 Case studies

We apply the neural network machinery to detect the direction of the time's arrow and compare the output of the network with the theoretical optimal result of Eq. (4.2). We first consider a single Brownian particle in a moving potential and then move on to the more complicated problem of a spin chain with nearest-neighbour coupling in a magnetic field and discuss how controlling the field and the coupling affect the results. The networks not only learn to guess the direction of the time's arrow but also learn to closely reproduce the likelihood function.

4.4.1 Brownian particle in a moving potential

An overdamped Brownian particle at temperature β^{-1} in a harmonic potential (Fig. 4.2(a)), evolves according to

$$\dot{x} = -\frac{k}{\gamma}(x - \lambda) + \xi(t), \tag{4.9}$$

where k denotes the strength of the potential, λ is the position of the center of the potential, and γ is the damping rate. The noise term $\xi(t)$ satisfies $\langle \xi(t)\xi(t') \rangle = 2(\beta\gamma)^{-1}\delta(t - t')$. In the forward protocol, the value of λ is changed from A to B at a fixed rate $\dot{\lambda} = u$. Hence the reverse protocol changes λ from B to A with $\dot{\lambda} = -u$.

After generating samples of the forward and backward trajectories using Eq. (4.9) (Figs. 4.2(b) and (c)), we train a classifier to predict the label for a given trajectory, as described earlier. In this example it is easy to detect the direction of the time's arrow, as the work distributions have a modest overlap, see Fig. 4.2(d). In Fig. 4.2(e) we compare the accuracy and the output of a LR classifier (grey dots) with the theoretical likelihood (solid curve) obtained from Eq. (4.2). The remarkable agreement with the theory can be understood by noting that work W, calculated by numerically integrating $\dot{W} = -ku(x - ut)$ for a given trajectory, is a linear function of the elements of the trajectory matrix $\mathbf{X}$. Therefore, as we show in Section 4.6 LR is well-equipped to calculate this quantity and reproduce the likelihood function.

4.4.2 Spin chain in a magnetic field

Now let us consider a more complicated, many-particle system and a non-linear work protocol, namely a spin chain in a magnetic field, in contact with a thermal reservoir at temperature β^{-1}, described by a Hamiltonian

$$H = J(t) \sum_i \sigma_i \sigma_{i+1} - B(t) \sum_i \sigma_i, \tag{4.10}$$

where $\sigma_i \in \{-1, +1\}$ is the spin variable at site i, $J(t)$ is the nearest-neighbour coupling strength, and $B(t)$ is the magnetic field. The dynamics of this system are modeled as a Markov process. The Hamiltonian aligns the spin in preferred energy configurations, while thermal fluctuations cause the spins to flip randomly. We consider two scenarios. First, the coupling is assumed to be constant, and the

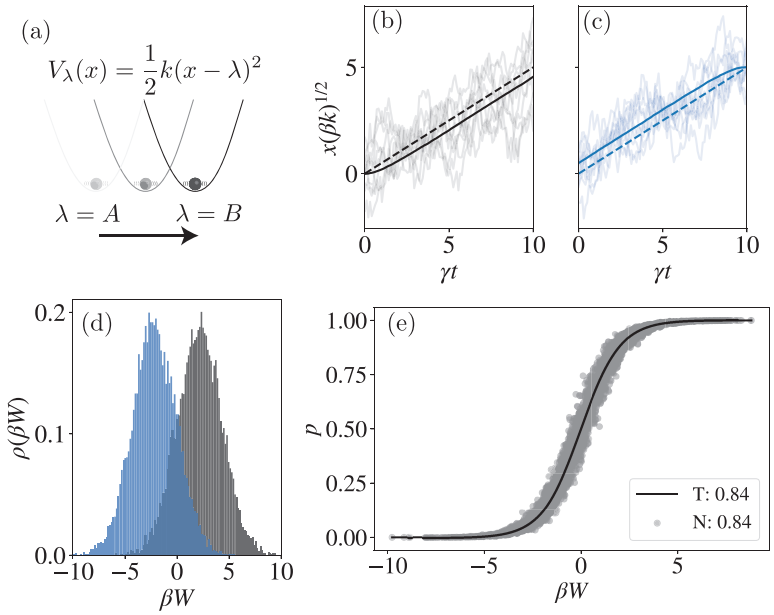

Figure 4.2: Brownian particle in a moving potential. (a) An overdamped Brownian particle with damping rate γ at temperature β^{-1} is in a harmonic potential $V_\lambda(x)$, with stiffness k. The position of the potential's center, λ, is externally controlled and is moved from A to B in the forward process. (b) Sample trajectories (grey) and the average trajectory (black) in the forward protocol. Note that the average trajectory lags behind the center of the potential (dashed line). In the simulations, γ, β, k are chosen to be 1, and $\lambda = 0$ at $t = 0$ is varied to $\lambda = 5$ at $t = 10$. (c) Sample backward trajectories (light blue) and their average (dark blue). The average trajectory leads the potential's center (dashed lines). (d) Work distribution for the forward (black) and the backward (blue) trajectories. They are both normally distributed and are symmetric around 0. (e) The likelihood of the forward process for a set of test trajectories. The output of the neural network (N), p, over the test set (grey dots) resembles the theoretical (T) likelihood $P(\mathrm{F}|\mathbf{X})$ (solid black line). The numbers in legends denote the accuracy of the methods over 4000 test trajectories. The classifier is trained on 12000 samples.

magnetic field is varied in time, see Fig. 4.3(a). Next, the magnetic field is constant and the coupling is varied, see Extended Data Fig. 1. We refer to the former as the $\mathcal{B}$-protocol, and the latter as the $\mathcal{J}$-protocol.

In the forward process of the $\mathcal{B}$-protocol, $J(t)$ is constant over time, and $B(t) = B_0 \cos(\pi t/\tau)$ changes from $B_0 > 0$ at $t = 0$ to $-B_0$ at $t = \tau$, as shown in Fig. 4.3(b). At low temperatures and large magnetic fields, the spins are aligned

with the direction of the field. As the temperature is increased, fluctuations become more prominent (see Fig. 4.3(c) and (d)). These fluctuations increase the overlap in work distributions, blurring the direction of time's arrow, see Fig. 4.3 (e). We train a single classifier using samples of forward and backward trajectories for three different temperatures generated by the Metropolis algorithm. This training scheme, known as multi-task learning, improves the performance and generalizability of the classifier [102]. We observe that the success of LR in learning both the correct labels and in approximating the likelihood function persists, see Fig. 4.3(f). The reason, again, lies in the functional form of W, which is evaluated by numerically integrating $\dot{W} = -\dot{B}(t)\sum_i \sigma_i$, and thus is proportional to a weighted sum of the elements of the input trajectory. Since $\Delta F = 0$ in this protocol, LR is a perfect model of the likelihood function for all the temperatures.

The $\mathcal{J}$-protocol is more complicated and has a ferromagnetic-antiferromagnetic transition. In this protocol, $B(t)$ is constant in time and $J(t) = J_0 \cos(\pi t/\tau)$ is varied non-linearly from $J_0 > 0$ at $t = 0$ to a $-J_0$ at $t = \tau$ in the forward process. The LR classifier does not perform well in this case, see Figure 4.9, as $\dot{W} = \dot{J}(t)\sum_i \sigma_i\sigma_{i+1}$ is no longer linearly related to the input. However, by using a CNN with periodic boundary condition we are able to recover the optimal accuracy and obtain results similar to those for the $\mathcal{B}$-protocol (see Fig. 4.4). The convolution layer in a CNN has filters that can capture the two-body nearest-neighbor correlations required to calculate the work (see Section 4.6.3). Note that in this process $\Delta F \neq 0$, which adds to the complexity of the problem.

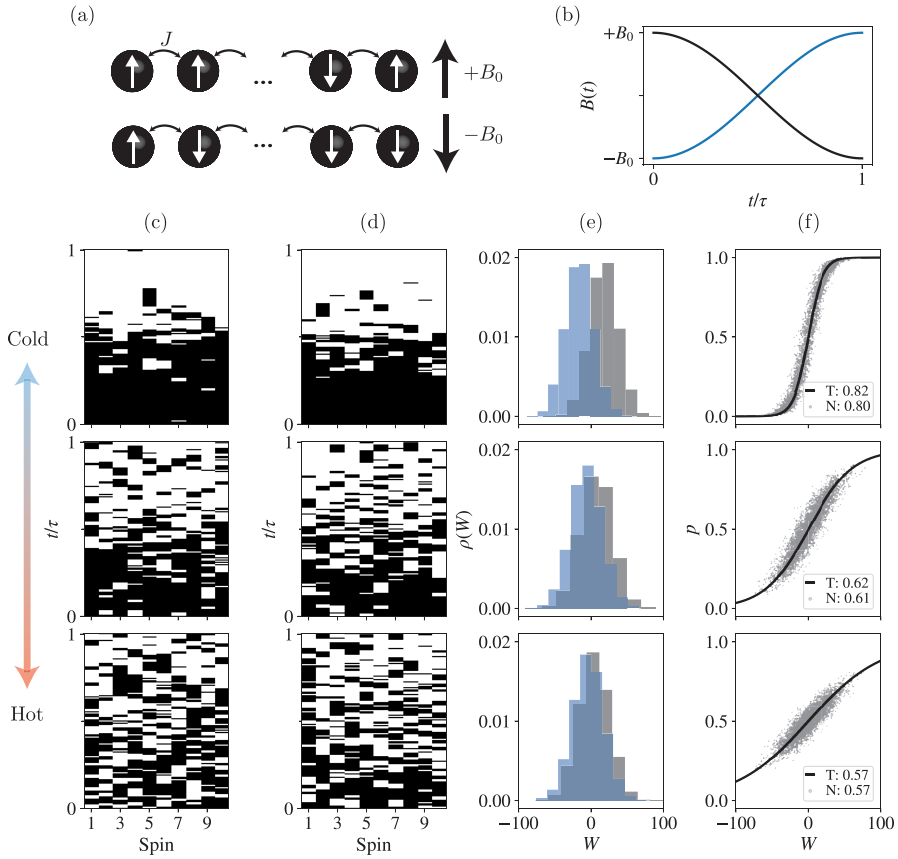

Figure 4.3: Spin chain in a time-dependent magnetic field. (a) A chain of ten spins with periodic boundary condition is placed in a magnetic field. The strength of coupling between nearest neighbors $J(t) = -1$ is fixed, and all the quantities are in the units of $|J|$. The forward process starts with spins in equilibrium at temperature β^{-1} with $B(0) = +B_0 > 0$ at time $t = 0$ and ends at a non-equilibrium state with $B(\tau) = -B_0$ at at time $t = \tau$. (b) The forward (black) and the reverse (blue) protocols $B(t)$. In our examples $B_0 = 20$ and $\tau = 500$. (c) Sample forward trajectories encoded in a matrix, where the black and white pixels (± 1 entries) denote spins pointing up and down, respectively. The rows and columns correspond to time steps and spin positions, respectively. (d) Sample backward trajectories obtained from time-reversal of the reverse process. (e) The distribution of work $\rho(W)$ for the forward (black) and backward (blue) trajectories, (f) the theoretical likelihood function (solid black line) and the output of the neural network p over the test set (grey dots) for various temperatures. The numbers in legends denote the accuracy of the theory (T) and neural network (N) over 4000 test trajectories for each temperature. In this example, a single classifier is trained simultaneously with sample trajectory data with three different temperatures (12000 samples for each temperature). The temperatures corresponding to different rows in panels (c), (d), (e), and (f) are $\beta^{-1} = 10, 30, 50$ in descending order. As the temperature increases, the distinction between the forward and backward trajectories is blurred.

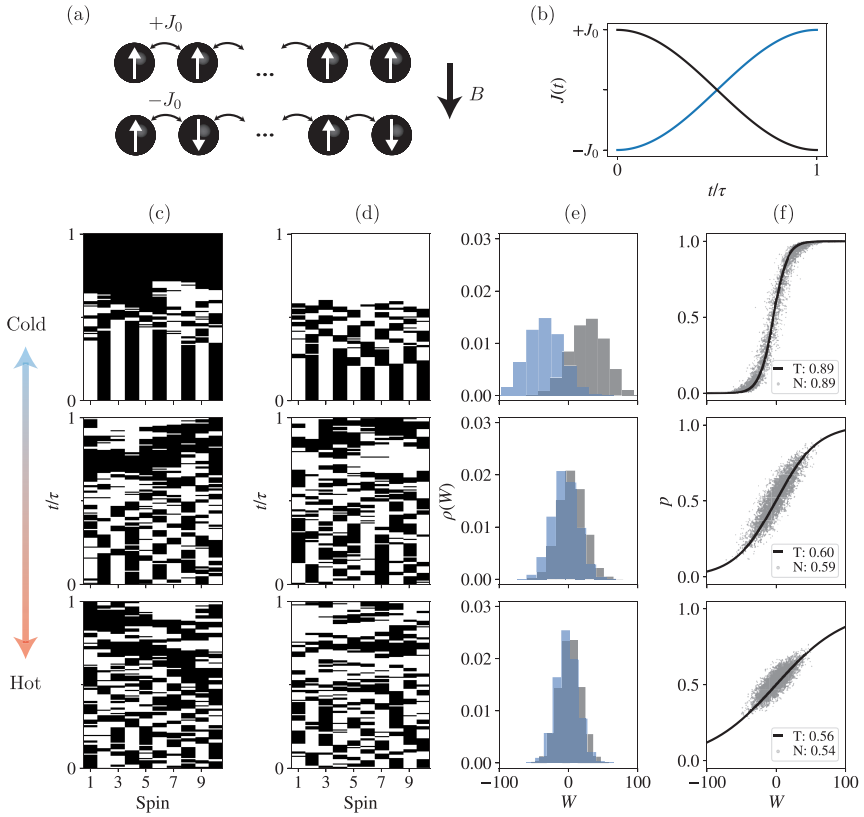

Figure 4.4: Spin chain with a time-dependent coupling. (a) A chain of ten spins with periodic boundary condition is placed in a constant magnetic field $B(t) = -1$, and all the quantities are in the units of $|B|$. The time-dependent coupling between nearest neighbors is $J(t)$. The forward process starts with spins in equilibrium at temperature β^{-1} with $J(0) = +J_0 > 0$ and ends at a non-equilibrium state with $J(\tau) = -J_0$. (b) The forward (black) and the reverse (blue) protocols $J(t)$. In our examples $J_0 = 20$ and $\tau = 500$. (c) Sample forward trajectories encoded in a matrix, where the black and white pixels (± 1 entries) denote spins pointing up and down, respectively. The rows and columns correspond to time steps and spin positions, respectively. (d) Sample backward trajectories obtained from time-reversal of the reverse process. (e) The distribution of work $\rho(W)$ for the forward (black) and backward (blue) trajectories, (f) the theoretical likelihood function (solid black line) and the output of the neural network p over the test set (grey circles) for various temperatures. The numbers in legends denote the accuracy of the theory (T) and logistic regression (NN) over 4000 test trajectories for each temperature. In this example, a single classifier is trained simultaneously with sample trajectory data with three different temperatures (12000 samples for each temperature). The temperatures corresponding to different rows in panels (c) - (f) $\beta^{-1} = 10, 30, 50$ in descending order. As the temperature increases, the distinction between the forward and backward trajectories is blurred.

4.5 Interpretation and extensions

We now use four approaches to investigate trained networks and to develop insight into what they have learned.

4.5.1 Inceptionism and dreaming

First, we use inceptionism techniques [103, 104] to learn the network's ideal representative of forward and backward trajectories. Specifically, we use a Monte Carlo approach to transform a randomly selected input trajectory into one for which the networks trained for $\mathcal{B}$ and $\mathcal{J}$-protocols output 1 or 0 corresponding to forward and backward trajectories, respectively. This is in contrast with the previous section where we optimized for the weights and biases of the network. Among the simulated trajectories in the test set, we choose one with $p \approx 0.5$ [104] – this is a trajectory for which the network has difficulty assigning the direction of time's arrow. We propose random spin flips and accept those moves that cause the output to get closer to the desired value of 1 or 0. We additionally demand that there be at most 1 spin-flip per time step, to ensure that the network 'dreams' of trajectories consistent with our simulations. We find that the networks' ideas of the forward and backward trajectories show strong agreement with the true physical picture, see Fig. 4.5(a).

4.5.2 Coarse-grained features and the reduced phase-space

To assign a physical interpretation to the networks' decision-making process, we project the trajectories onto a two-dimensional reduced phase space cor-

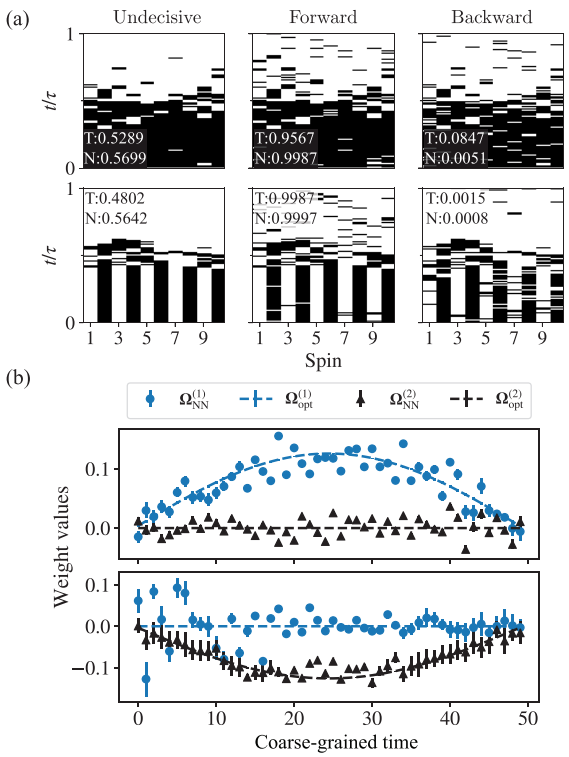

Figure 4.5: Interpreting the neural network's inner mechanism. (a) Starting with a random trajectory (leftmost column), we ask the trained network to 'dream' of its idea of the forward (middle column) and backward (rightmost column) trajectories. Here, t is time and τ is the duration of the process. The top row corresponds to the $\mathcal{B}$-protocol, with parameters identical to those in Fig. 4.3, and the bottom row corresponds to the $\mathcal{J}$-protocol with similar parameters, but with values of magnetic field $B(t)$ and the coupling strength $J(t)$ interchanged. The black and white pixels denote spins pointing up and down, respectively. The numbers in the inset indicate the forward likelihood $P(\mathrm{F}|\mathbf{X})$, obtained from the theory (T) using Eq. (4.2) and from the neural network's output (N). (b) The weights of the networks trained using coarse-grained trajectories associated with the magnetization $\mathbf{\Omega}^{(1)}$ and the nearest-neighbour correlations $\mathbf{\Omega}^{(2)}$. The network for the $\mathcal{B}$-protocol (top row) is trained simultaneously at temperatures $\beta^{-1} = 10, 30, 50$ and the network for the $\mathcal{J}$-protocol (bottom row) is trained at $\beta^{-1} = 10$. The x-axis represents coarse-grained time of 10 steps in trajectories in panel (a). The error bars are standard deviation over 10 trained networks with random weight initialization. The network bases its decision on the magnetization in the former, and on the nearest-neighbor correlations in the latter case. If the values of $\mathbf{\Omega}^{(\ell)}_{\mathrm{NN}}$s of the trained networks (markers) match the optimal weights $\mathbf{\Omega}^{(\ell)}_{\mathrm{opt}}$ (dashed line), the output of the network agrees with the exact likelihood (4.2).

responding to the collective coordinates $\{\tilde{x}^{(1)}(t)\} = \{\sum_i \sigma_i(t)\}$ and $\{\tilde{x}^{(2)}(t)\} = \{\sum_i \sigma_i(t)\sigma_{i+1}(t)\}$ (taking period boundary conditions), representing magnetization

and nearest-neighbour correlations, respectively. We also replace the value of $\tilde{x}^{(\ell)}(t)$ within each time window of ten time steps, by the sum of the values within that window. Specifically, for the two protocols concerning the spin chain in a magnetic field, the coarse-grained features are

$$\tilde{x}_s^{(1)} = \sum_{t=ms}^{m(s+1)-1} \sum_{i=1}^{n} X_{t,i}, \tag{4.11}$$

$$\tilde{x}_s^{(2)} = \sum_{t=ms}^{m(s+1)-1} \sum_{i=1}^{n} X_{t,i} X_{t,i+1}, \tag{4.12}$$

where m is an integer and s is the scaled time. We use this feature map as the input to a LR classifier. Note that the input to the network is a $2\tau/m$ dimensional vector $\begin{bmatrix} \tilde{\mathbf{x}}^{(1)} \\ \tilde{\mathbf{x}}^{(2)} \end{bmatrix}$, and the weights corresponding to $\tilde{\mathbf{x}}^{(\ell)}$ are denoted by τ/m dimensional vectors $\mathbf{\Omega}^{(\ell)}$ for $\ell = 1, 2$.

By thus coarse-graining in both phase space and time, we reduce the noise due to finite size effects and variations over samples. We use these coarse-grained trajectories to train LR classifiers for both $\mathcal{B}$ and $\mathcal{J}$-protocols (See Figure 4.10 for the performance of these classifiers). Finally, we investigate the weights $\mathbf{\Omega}^{(\ell)}$ that the networks assign to the magnetization ($\ell = 1$) and the nearest neighbor correlations ($\ell = 2$) of an input trajectory. Fig. 4.5(b) reveals that for the $\mathcal{B}$-protocol (top row), the network mostly cares about the magnetization, whereas when the $\mathcal{J}$-protocol is performed (bottom row), the network bases its decision on the nearest-neighbor correlations. Moreover, the learned values of $\mathbf{\Omega}^{(\ell)}$ agree with our analytical results that reproduces the correct likelihood value (see also Sec. 4.6.4).

These observations suggest that the network learns that the time derivative of the Hamiltonian, and by extension the work (4.3), is an important feature in guessing the direction of time's arrow. We note that when the process is highly irreversible, the distributions of the forward and reverse work are well-separated. In such cases, we observe that while LR has an almost 100% accuracy, it does not learn work (obtained by inverting the sigmoid function in the output). This is because the events that enable the network to learn work are extremely rare, and are usually absent in the training data. However, there are other evident differences that can show the direction of time's arrow. We show an example of such a process for the $\mathcal{B}$-protocol at low temperatures. We observe that the orientation of spins undergoes a sharp transition as the magnetic field changes sign. However, the time during which this transition occurs is different in the forward and backward trajectories. The classifier makes a decision based on the spin configuration at this particular time. See Figure 4.6.

4.5.3 Mixture of experts

By using our knowledge about the structure of the problem we design a neural network that can learn to accurately guess the direction of time's arrow for trajectories generated using multiple protocols, when the identity of the protocol is not specified. Specifically, we use a mixture of experts (MoE), with an output that is the weighted sum of expert networks [105]. When the protocol is not specified, the

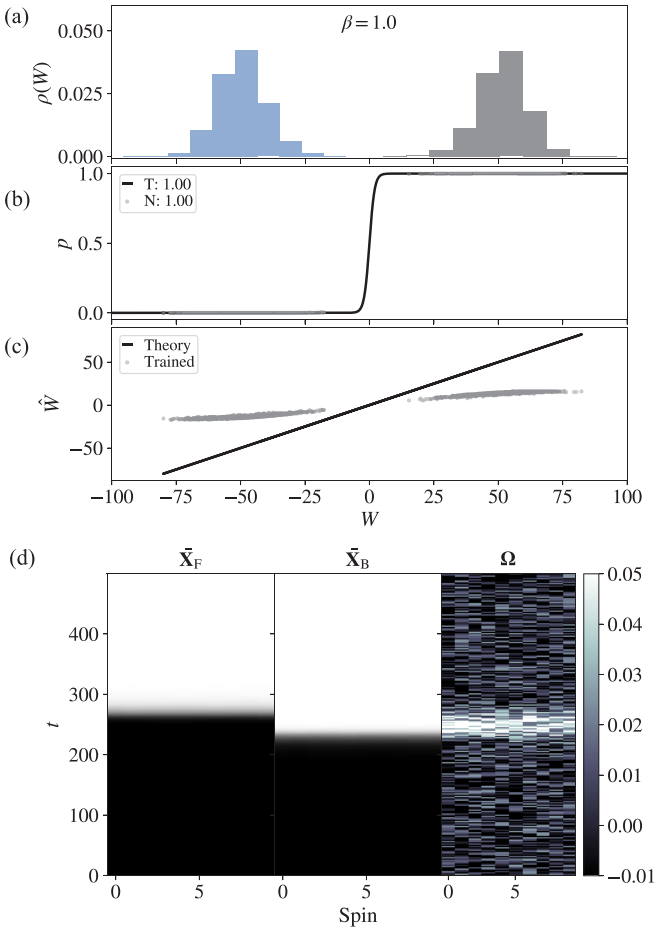

Figure 4.6: A highly irreversible version of the $\mathcal{B}$-protocol, with parameters similar to those in Figure 3 in the main text at a lower temperature $\beta^{-1} = 1$. All in the qunatities are in the units of the fixed coupling strength $|J|$. (a) The distributions of forward (grey) and reverse (blue) work $\rho(W)$ are well-separated. (b) The forward likelihood of sample trajectories is either 0 or 1, and the classifier's prediction (grey dots), matches the theory (solid curve). The numbers in the legend denote the accuracy of the theory (T) and the neural network (N). (c) In this example, the value of work that the network calculates $\hat{W}$ (obtained by inverting the sigmoid function) is different than the actual value of work W. (d) The average forward ($\bar{\mathbf{X}}_\mathrm{F}$) and backward ($\bar{\mathbf{X}}_\mathrm{B}$) trajectories, and the network wights $\mathbf{\Omega}$ over time t suggests that the spin orientations midway through the process is a way to decide the direction of time's arrow.

net forward likelihood is

$$P(\mathrm{F}|\mathbf{X}) = P(\mathrm{F}|\mathbf{X}, \mathcal{B})P(\mathcal{B}|\mathbf{X}) + P(\mathrm{F}|\mathbf{X}, \mathcal{J})P(\mathcal{J}|\mathbf{X}). \tag{4.13}$$

The quantities $P(F|\mathbf{X}, \mathcal{B})$ and $P(F|\mathbf{X}, \mathcal{J})$ are modeled using neural networks similar to those previously considered for $\mathcal{B}$ and $\mathcal{J}$-protocols, respectively. These networks are referred to as experts. Additionally, we use a CNN to model $P(\mathcal{B}|\mathbf{X}) = 1 - P(\mathcal{J}|\mathbf{X})$. This CNN, which is called the gating network, learns the protocol from trajectories. Therefore, we obtain a larger three-headed network by combining the output of the three neural networks as in Eq. (4.13), as illustrated in Fig. 4.7(a). For the training, we use the pre-trained expert networks for the $\mathcal{B}$ and $\mathcal{J}$-protocols, and optimize the cross entropy cost function over sample trajectories from both protocols. We observe that the performance of this network is similar to that of the individual networks, as the gating network learns to accurately identify the protocol of input trajectories (see Fig. 4.7(b)). Note that the predictions of the gating network are more accurate at lower temperatures. This makes sense as the distribution of the initial state in the two protocols are distinguishable in low temperatures, but become less so as the temperature is increased.

4.5.4 Alternative activation functions

The logistic function that appears in the theoretically calculated likelihood in the time's arrow problem, is similar to the activation function that is commonly used for classification in machine learning. To assess the general ability of the networks in approximating the likelihoods we try different activation functions. Specifically, we choose $g(z) = \exp(-z^2)$ and $g(z) = \sin^2(z)$ as the activation of the last layer of the neural network so that the output is always between 0 and 1, and can be

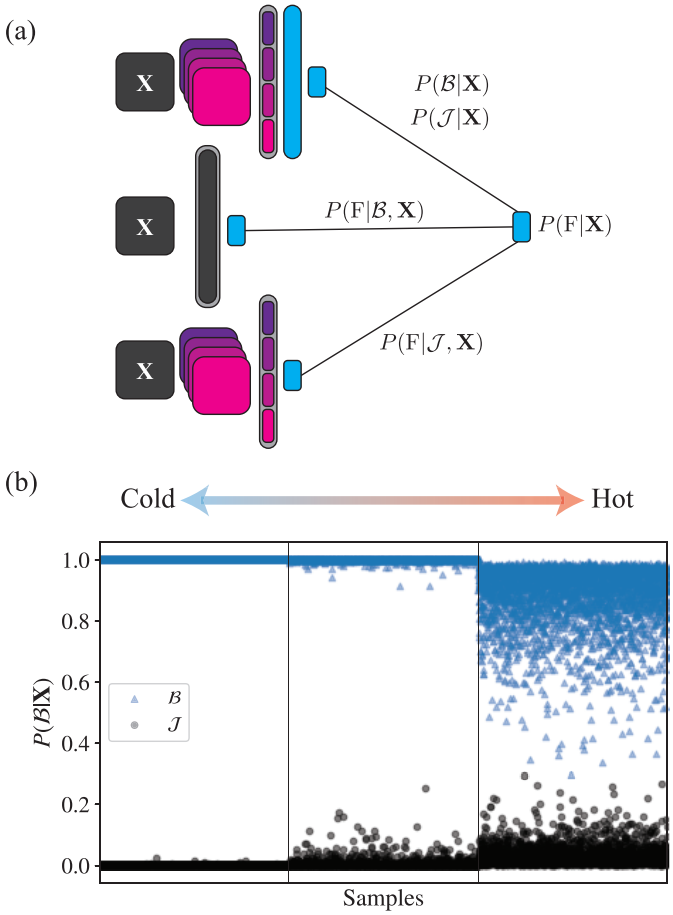

Figure 4.7: Mixture of experts (MoE). (a) The MoE network models the forward likelihood $P(\mathrm{F}|\mathbf{X})$ for an input trajectory $\mathbf{X}$. It consists of a gating CNN that predicts the protocol $P(\mathcal{B}(\mathcal{J})|\mathbf{X})$, and two networks that predict the forward likelihood of a trajectory given the protocol $P(\mathrm{F}|\mathcal{B}(\mathcal{J}),\mathbf{X})$. (b) The output of the gating network, which models $P(\mathcal{B}|\mathbf{X})$, is shown for different sample trajectories of the $\mathcal{B}$ (blue triangles) and $\mathcal{J}$ (black dots) protocols. The trajectories for the $\mathcal{B}$-protocol correspond to the same parameters used in Fig. 4.3 with three temperatures $\beta^{-1} = 10, 30, 50$ in the units of $|J|$. The parameters in the $\mathcal{J}$-protocol are similar, but the values of magnetic field $B(t)$ and the coupling strength $J(t)$ are interchanged. The horizontal axis shows different samples in three temperature regions separated by vertical line. It is harder to predict the protocol at higher temperatures.

interpreted as probabilities. We also add a hidden layer to give the network the ability to calculate complex functions. We compare the network's output with the theoretical likelihoods for the spin chain under the $\mathcal{B}$-protocol with coarse-grained features discussed in Section 4.6. We only train the network at a single temperature

and observe that the network can still approximate the likelihood function as shown in Figure 4.8.

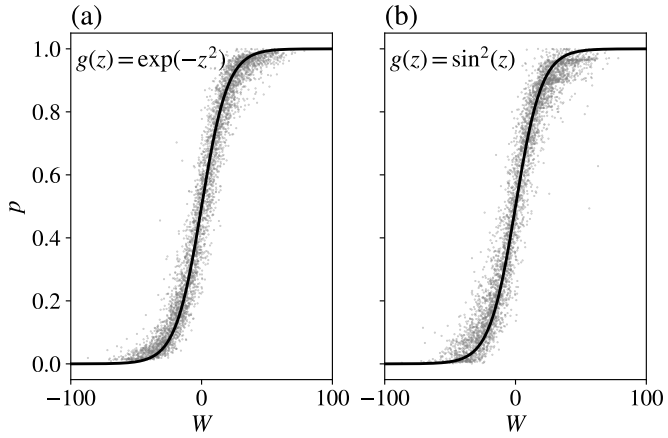

Figure 4.8: The theoretical likelihood function (solid black line) and the output of the neural network with custom activation functions (grey dots) for the last layer corresponding to the coarse-grained $\mathcal{B}$-protocol at temperature $\beta^{-1} = 10$. The network has a hidden layer with 50 neurons with tanh activation. The last layer's activation functions are (a) $g(z) = \exp(-z^2)$ and (b)$g(z) = \sin^2(z)$. The accuracy of both models is 0.80, which is close to the theoretically optimal value of 0.82. Comparing the plots with the leftmost column of Figure 4.10b, we observe that the performance slightly deteriorates. However, the networks still capture the essence of the likelihood function.

4.6 Optimal Networks

For some of the examples that we considered, it is possible to derive analytical expressions for the optimal weights and biases of the network. Specifically, we examine the expression that is used to calculate the work W and the change in free energy ΔF. Because the logistic sigmoid activation function, i.e., $g(z) = 1/(1 + \exp(-z))$, used for classification coincides with the form of the likelihood function $P(\mathrm{F}|\{\mathbf{x}(t)\}) = 1/(1 + e^{-\beta(W-\Delta F)})$, in the arrow of time problem [17], we are able to find the networks parameters $\{\mathbf{\Omega}, \mathbf{b}\}$ that reproduce the same likelihood function.

To illustrate this, consider the LR model with the output $p = 1/(1 + \exp(-z))$ and $z = \mathbf{\Omega}^\intercal \mathbf{a} + b$. If we find $\mathbf{\Omega}$ and b such that $z = \beta(W - \Delta F)$, the output of the network p correctly represent $P(\mathrm{F}|\mathbf{X})$. In the following we show that when W is linear in the elements of $\mathbf{X}$, and therefore linear in the elements of $\mathbf{a} = \mathrm{vec}(\mathbf{X})$, we are able to find such optimal $\mathbf{\Omega}$ and b.

4.6.1 Brownian particle in a moving potential

In this example, the system's state at each time step is described by a scalar x_t, as the position of the particle. We have a total of $\tau + 1$ time steps, therefore, the input to the NN is a $\tau + 1$ dimensional vector. The LR classifier considered here is parameterized by a $\tau + 1$ dimensional weight vector with elements Ω_t for $t = 0, \ldots, \tau$ and a bias b. We find

$$W = \sum_{t=0}^{\tau-1} \frac{1}{2} k\delta\lambda(-2x_t + \lambda_t + \lambda_{t+1}), \tag{4.14}$$

where $\delta\lambda = \lambda_{t+1} - \lambda_t$ is independent of t, because the protocol is linear. Note that $\Delta F = 0$ in this example. With the choice of

$$\Omega_t = \begin{cases} -\beta k\delta\lambda & (t \neq \tau) \\ 0 & (t = \tau) \end{cases}, \tag{4.15}$$

$$b = \sum_{t=0}^{\tau-1} \frac{1}{2} \beta k\delta\lambda(\lambda_t + \lambda_{t+1}), \tag{4.16}$$

we can see that $\mathbf{\Omega}^\intercal \mathbf{a} + b = \beta(W - \Delta F)$, where $(\mathbf{a})_t = x_t$.

4.6.2 Spin chain - $\mathcal{B}$-protocol

The full trajectory of an n-spin system over τ time steps is represented by a $\tau \times n$ matrix $\mathbf{X}$. We denote the orientation (up or down) of the ith spin at time t with $X_{t,i} = \pm 1$. The input to LR classifier, is a vector obtained from rearranging the trajectory matrix $\mathbf{X}$ to shape it into an $\tau n \times 1$ array. We find that

$$W = -\sum_{t=0}^{\tau-1}(\delta B_t \sum_{i=1}^{n} X_{t,i}), \tag{4.17}$$

where $\delta B_t = B_{t+1} - B_t$. Work calculated using Eq. (4.17) is the discrete time version of W obtained from $\dot{W} = -\dot{B}(t)\sum_i \sigma_i$. In this example, $\Delta F = 0$ again, and we find that the optimal weights and bias are given by

$$\Omega_{t,i} = -\beta \delta B_t \qquad (t \neq \tau), \tag{4.18}$$

$$b = 0, \tag{4.19}$$

where $\Omega_{\tau,i} = 0$. Note that if the input $X_{t,i}$ is scaled with β, a single LR classifier is able to reproduce the correct likelihood for different temperatures.

4.6.3 Spin chain - $\mathcal{J}$-protocol

Using the same notation in the previous section, we find that the work, W, is given by

$$W = \sum_{t=0}^{\tau-1}(\delta J_t \sum_{i=1}^{n} X_{t,i} X_{t,i+1}), \tag{4.20}$$

where $\delta J_t = J_{t+1} - J_t$. Work calculated using Eq. (4.20) is the discrete time version of W obtained from $\dot{W} = \dot{J}(t) \sum_i \sigma_i \sigma_{i+1}$. We also use periodic boundary condition, which implies $X_{t,n+1} = X_{t,1}$. Moreover, there is a non-zero change in the free energy, which is given by [106]

$$\Delta F = -\frac{1}{\beta} \log\left(\frac{\epsilon_-(\beta B, \beta J_0)^n + \epsilon_+(\beta B, \beta J_0)^n}{\epsilon_-(\beta B, -\beta J_0)^n + \epsilon_+(\beta B, -\beta J_0)^n}\right) \tag{4.21}$$

where

$$\begin{aligned} \epsilon_\pm(\beta B, \beta J) = \exp(\beta J)\cosh(\beta B) \\ \pm \sqrt{\exp(2\beta J)\cosh^2(\beta B) - 2\sinh(2\beta J)}. \end{aligned} \tag{4.22}$$

We see that it is not possible to have a logistic regression model that calculates W accurately, as it is a non-linear function of the elements of $\mathbf{X}$ (see Figure 4.9).

However, a convolutional neural network (CNN) can learn the relevant representation (i.e. nearest-neighbor correlations) from the input data, and learn the corresponding weights to calculate work. Specifically, to show that in principle a CNN can exactly calculate the correct likelihood from the input, we consider a CNN with four 1×2 filters with periodic boundary condition and the rectifier activation g_1, followed by the output layer with sigmoid activation g_2. We set the biases of the convolutional layer to zero, and choose the weights $\mathbf{\Omega}^{[1,i]}$ for filters $i = 1, 2, 3, 4$ as

follows

$$\begin{Bmatrix} \mathbf{\Omega}^{[1,1]} \\ \mathbf{\Omega}^{[1,2]} \\ \mathbf{\Omega}^{[1,3]} \\ \mathbf{\Omega}^{[1,4]} \end{Bmatrix} = \begin{Bmatrix} (1 & 1) \\ (1 & -1) \\ (-1 & 1) \\ (-1 & -1) \end{Bmatrix}. \tag{4.23}$$

Each filter is only activated for one of the possible configuration of two neighboring spins. Specifically, given the $\tau \times n$ input $\mathbf{X}$, the output of each filter $g_1(\mathbf{\Omega}^{[1,j]} * \mathbf{X})$ is a $\tau \times n$ matrix $\tilde{\mathbf{X}}^{(j)}$ such that

$$\tilde{X}^{(1)}_{t,i} = 1 \quad \text{if} \quad (X_{t,i}X_{t,i+1}) = (1,1), \tag{4.24}$$

$$\tilde{X}^{(2)}_{t,i} = 1 \quad \text{if} \quad (X_{t,i}X_{t,i+1}) = (1,-1), \tag{4.25}$$

$$\tilde{X}^{(3)}_{t,i} = 1 \quad \text{if} \quad (X_{t,i}X_{t,i+1}) = (-1,1), \tag{4.26}$$

$$\tilde{X}^{(4)}_{t,i} = 1 \quad \text{if} \quad (X_{t,i}X_{t,i+1}) = (-1,-1), \tag{4.27}$$

and $\tilde{X}^{(j)}_{t,i} = 0$ otherwise. We can now rewrite the output of the network as

$$g_2(b_2 + \sum_j (\sum_{t,i} \Omega^{[2,j]}_{t,i} \tilde{X}^{(j)}_{t,i}))), \tag{4.28}$$

where $\mathbf{\Omega}^{[2,j]}$ contains the weights of the output layer corresponding to $\tilde{\mathbf{X}}^{(j)}$. The

optimal values of these weights are given by

$$\Omega^{[2,1]}_{t,i} = \Omega^{[2,4]}_{t,i} = -\beta\delta J_t \qquad (t \neq \tau), \tag{4.29}$$

$$\Omega^{[2,2]}_{t,i} = \Omega^{[2,3]}_{t,i} = +\beta\delta J_t \qquad (t \neq \tau), \tag{4.30}$$

$$b = -\beta\Delta F, \tag{4.31}$$

where $\Omega^{[2,j]}_{\tau,i} = 0$.

Therefore, a CNN with four 1×2 filters is sufficient to capture the likelihood at a single temperature of β^{-1}. We find that in practice, a CNN with such an architecture is likely to get stuck at local minima, and finding the optimal parameters shown above greatly depends on the initial weights of the network. However, we observe that a CNN with four 2×2 filters can achieve a close to optimal performance more easily.

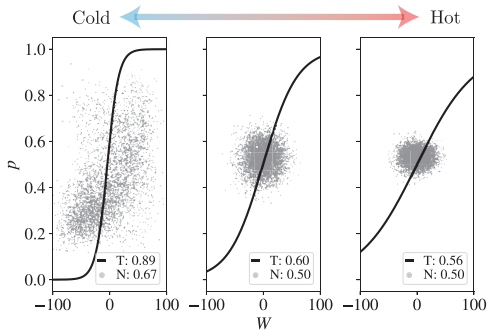

Figure 4.9: Performance of logistic regression in the $\mathcal{J}$-protocol. In the $\mathcal{J}$-protocol, the magnetic field $B(t) = -1$ is constant, and all the quantities are expressed in the units of $|B|$. The coupling strength varies as $J(t) = J_0 \cos(\pi t/\tau)$ with $J_0 = 20$ and $\tau = 500$. The LR classifier is trained simultaneously with sample trajectory data with three different temperatures $\beta^{-1} = 10, 30, 50$, corresponding to the columns in the figure from left to right, respectively. We use 12000 samples for each temperature in the training. The theoretical likelihood function (solid black line) and the output of the neural network p over the test set (grey dots) are plotted for various temperatures as a function of the work performed on the system W. LR does not perform well and cannot match the performance of a CNN as observed in Fig. 4 in the main text. The numbers in legends denote the accuracy of the theory (T) and neural network (N) over 4000 test trajectories for each temperature.

4.6.4 Coarse-grained features

To obtain Figure 4.5(b) in the main text, the trajectories are coarse-grained according to

$$\tilde{x}_s^{(1)} = \sum_{t=ms}^{m(s+1)-1} \sum_{i=1}^{n} X_{t,i}, \tag{4.32}$$

$$\tilde{x}_s^{(2)} = \sum_{t=ms}^{m(s+1)-1} \sum_{i=1}^{n} X_{t,i} X_{t,i+1}, \tag{4.33}$$

where m is an integer and s is the scaled time. Using this feature map, LR classifier can calculate W for both $\mathcal{B}$ and $\mathcal{J}$-protocols (See Figure 4.10). The input to the

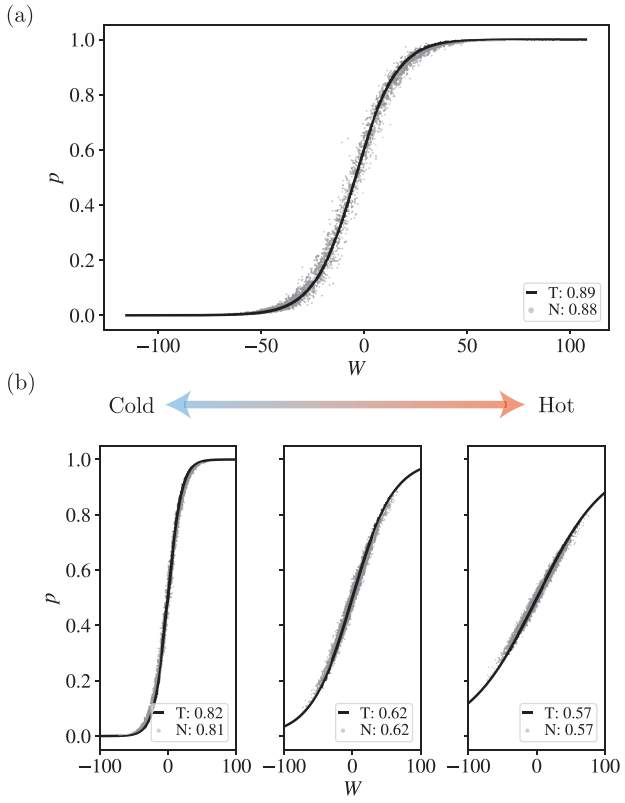

Figure 4.10: Prediction of logistic regression with coarse-grained features. The output of the network (grey dots) and the theoretical likelihood function (solid curve) agree remarkably for both $\mathcal{B}$ and $\mathcal{J}$-protocols. Coarse-graining and feature engineering improves the performance. The parameters of the protocols and the units are similar to those used in Figs. 3 and 4 in the main text for $\mathcal{B}$-protocol and $\mathcal{J}$-protocol, respectively. (a) LR's prediction trained and tested at temperature $\beta^{-1} = 10$ as a function of the work performed on the system W . (b) LR predicting the direction of time's arrow in the $\mathcal{B}$-protocol as a function of W, trained and tested simultaneously for three different temperatures $\beta^{-1} = 10, 30, 50$, respectively.

network is a $2\tau/m$ dimensional vector $\begin{bmatrix} \tilde{\mathbf{x}}^{(1)} \\ \\ \tilde{\mathbf{x}}^{(2)} \end{bmatrix}$. We denote the weights corresponding to $\tilde{\mathbf{x}}^{(\ell)}$ by τ/m dimensional vectors $\mathbf{\Omega}^{(\ell)}$ for $\ell = 1, 2$. In this case, we approximate the optimal weights and bias of the networks by their average over the coarse-grained

time window. For the $\mathcal{B}$-protocol we find

$$\Omega_s^{(1)} = \frac{\beta}{m} \sum_{t=ms}^{m(s+1)-1} \delta B_t, \tag{4.34}$$

$$\Omega_s^{(2)} = 0, \tag{4.35}$$

$$b = 0. \tag{4.36}$$

Similarly, the weights and bias for the $\mathcal{J}$-protocol are given by

$$\Omega_s^{(1)} = 0, \tag{4.37}$$

$$\Omega_s^{(2)} = \frac{\beta}{m} \sum_{t=ms}^{m(s+1)-1} \delta J_t, \tag{4.38}$$

$$b = -\beta \Delta F. \tag{4.39}$$

In both cases we can see that $\beta(W - \Delta F) \approx (\mathbf{\Omega}^{(1)})^\intercal \tilde{\mathbf{x}}^{(1)} + (\mathbf{\Omega}^{(2)})^\intercal \tilde{\mathbf{x}}^{(2)}$, where the approximation comes from coarse-graining.

Note that in this case, even though LR classifier can calculate βW if the input is scaled with β, it is not possible to train the network over different temperatures. This is because $\Delta F \neq 0$, and a simple bias cannot capture multiple values of $\beta \Delta F$. Therefore, we only consider a single temperature for the $\mathcal{J}$-protocol in studying the optimal networks with coarse-grained features.

4.7 Conclusion and outlook

While in this Chapter we rediscovered elements of thermodynamics that were developed in recent decades, we are ultimately interested in answering open questions in thermodynamics. The physics of systems out-of-equilibrium is an area of interest, where there are unsolved questions that one may be able to answer by studying the dynamics of systems with machine learning algorithms. For instance, identifying the thermodynamic principle that determines the steady-state of a system [107], an outstanding problem in statistical physics, or deciding when a non-equilibrium system has an effective equilibrium description [108] are both examples where analyzing time-series data using machine learning could shed light on the inner mechanism of the problem.

Moreover, machine learning researchers have shown that ML techniques can be used to detect the playback direction of real-world videos [109, 110]. These studies are concerned with videos of macroscopic objects that are in principle irreversible, and the arrow of time has a clear direction. In such scenarios, there are many indicators that can reveal the true playback direction, and therefore it is hard to quantify the optimal performance. However, in the physical examples the optimal attainable accuracy of the classifier is dictated by the laws of physics. Therefore, problems with large number of phase-space coordinates and with complicated dynamics, such as the $\mathcal{J}$-protocol for 2D Ising model, can serve as a standardized benchmark for video classification algorithms.

Chapter 5: Conclusion and Outlook

In this dissertation, we examined characterization and control of open physical systems. We also considered applications of machine learning in scientific discovery. Throughout this work, we used our physical knowledge about the system that is being studied and combined it with advanced computational techniques to learn more about physical systems and improve their performance.

In the characterization front in Chapter 2 , we made assumptions about the nature of the noise and its temporal correlations, and designed a protocol to learn spatial correlations of the noise process using compressed sensing techniques. The temporal correlations of noise are typically studied under noise spectroscopy [5]. However, the type of operators and their spatial correlations are studied in quantum tomography and benchmarking, while assuming the noise is Markovian without any temporal signature [22]. Some benchmarking protocols, such as randomized benchmarking [111], can recognize the existence of non-Markovian noise, but they fall short of fully characterizing the temporal structure of the noise. Of course, fully characterizing a noise channel at different times without any assumptions about the model is costly. However, using the framework that we built and the techniques developed in our work, it might be possible to simultaneously characterize the spatial

and temporal correlations of certain noise processes.

Quantum experiments generate tons of data during every day calibration processes. Some of this data is unused as noted in Chapter 3, partly because the data is noisy or hard to model theoretically. We showed that a black box approach can take advantage of the information in the unused data and improve the performance of the system. While we showed an application of machine learning in the measurement process of trapped ion qubits, there is a great potential in applying machine learning techniques to other tasks in quantum information processing as shown in Refs. [112, 113]. Therefore, we can expect machine learning to play an import role in calibrating and automating quantum experiments [114, 115]. An important question that remains is the scalability of these techniques. In that regard, it would be interesting to see to what extent a verified quantum computer can help tuning another unverified quantum computer [116].

Finally, our ultimate goal is to study physics with the help of computer algorithms. In Chapter 5, we applied machine learning tools to a physical problem and gained insight about the underlying mechanism and the relevant physical parameters in the problem. We note that the theoretical understanding of the problem that we studied was completed in the past 20 years, whereas the algorithms utilized in the study were available decades before the development of those theories. It is conceivable that, with such numerical experiments, one could come up with the theoretical formulation of the solution before its discovery from physical principles. Therefore, interpretable machine learning techniques may help us learn physics from big data in the future. Techniques used in our work, combined with recently de-

veloped tools such as Neural SDEs [117] can also be used to study and model the dynamics of out-of-equilibrium systems. We believe our work could inspire future research directions in classical and quantum non-equilibrium systems as discussed in the following.

An outstanding question in non-equilibrium statistical physics is identifying the thermodynamics and the physical principles that determines the steady state of a system. Specifically, is there a principle similar to maximum entropy principle, that gives us a recipe for finding steady state of non-equilibrium systems? In other words, can we find a function of phase space variables, whose minimum/maximum with some constraints specify the steady state of the system? A concrete problem to study is self-oscillations of an electron shuttle [107]. The system has two modes of behavior and depending on the parameters of the problem, it spontaneously chooses one mode. While the dynamics of this problem is understood, it is not clear what physical principle determines this behavior. This is in contrast to problems in equilibrium physics, where we know that systems always try to minimize their free energy. Therefore, the ultimate question is finding the function that its optimum determines the steady state. To build the toolbox to answer this question we have to start with concrete examples that we know how to solve.

First, we can use equilibrium classical physics as a test bed and study what machines are able to learn about the system's equilibration process. Specifically, we can study the snapshots of a system's approach to equilibrium and try to learn that it is the free energy that is being minimized. We believe that this is achievable using techniques developed in inverse reinforcement learning and apprenticeship

learning [118]. By observing the nature, an algorithm can learn to mimic nature's behavior and take systems to equilibrium. An interesting question is whether the reward function learned by the algorithm coincides with the free energy function? After establishing this, we can apply our toolbox to the example of the electron-shuttle and more complicated problems such as Rayleigh–Bénard convection [119] and see what we can learn about the physics of the problem.

Additionally, there are interesting questions in the study of quantum systems out-of-equilibrium. For example, one interesting question is whether a driven-dissipative system has an effective equilibrium description [120, 121], or it falls under a genuine nonequilibrium universality class [122, 123, 124]. We believe that machine learning, and specifically time-series classification, can play an important role in identifying relevant features that lead to these distinct behaviors and shed light on underlying physical mechanisms. Another problem of interest is distinguishing chaotic versus non-chaotic, or integrable versus non-integrable systems. It is intriguing to ask whether one can identify relevant parameters that differentiates the two regimes, such as out-of-time-ordered correlators, by observing the system's dynamics.

Ultimately, we can repeat the question of what physical principle determines the approach to the steady state in quantum systems and try to identify the relevant features and the effective function that its minimum gives the steady state of the quantum system. Since quantum systems are inherently more complicated than their classical counterparts, we may need a quantum algorithm to achieve this, for example, by extending techniques developed in Ref. [125].

Appendix A: Optimal control for quantum detectors

This chapter is based on the author's contribution to Ref. [126]. We briefly describe the results and detail the author's contribution. We refer the reader to the original manuscript for more details [126].

A.1 Summary of the results

Quantum systems are promising candidates for sensing of weak signals as they can provide unrivaled performance when estimating parameters of external fields. However, when trying to detect weak signals that are hidden by background noise, the signal-to-noise-ratio is a more relevant metric than raw sensitivity. We identify, under modest assumptions about the statistical properties of the signal and noise, the optimal quantum control to detect an external signal in the presence of background noise using a quantum sensor. Interestingly, for white background noise, the optimal solution is the simple and well-known spin-locking control scheme. We further generalize, using numerical techniques, these results to the background noise being a correlated Lorentzian spectrum. We show that for increasing correlation time, pulse based sequences such as CPMG are also close to the optimal control for detecting the signal, with the crossover dependent on the signal frequency. These

results show that an optimal detection scheme can be easily implemented in near-term quantum sensors without the need for complicated pulse shaping.

A.2 Numerical optimization of control pulses

We consider a single qubit as a quantum sensor in the presence of a dephasing signal and noise. Assuming uniaxial control (along σ_x) and under the rotating wave approximation, the qubit Hamiltonian is given by,

$$H(t) = \frac{1}{2}J\left[\sqrt{\alpha}\ s(t) + \eta(t)\right]\hat{\sigma}_z + \frac{1}{2}\Omega(t)\hat{\sigma}_x, \tag{A.1}$$

where $\Omega(t)$ is the Rabi frequency of an arbitrary time-dependent control, the signal $s(t)$ and background noise $\eta(t)$ are both considered to be classical wide sense stationary Gaussian stochastic processes, α denotes the ratio of the signal-to-noise power (SNR) and J^2 is the total noise power. These stochastic processes have mean zero, $\overline{s(t)} = \overline{\eta(t)} = 0$ and two-point time correlations given by $\overline{\eta(t)\eta(t')} = g_\eta(t-t')$, $\overline{s(t)s(t')} = g_s(t-t')$ and $\overline{\eta(t)s(t')} = 0$ with the normalization $g_\eta(0) = g_s(0) = 1$ and $\overline{(\cdots)}$ denoting averaging over noise realizations. Alternatively, the noise correlations may also be represented in the frequency domain, using the power spectrum, $S_\eta(\omega) = \int_{-\infty}^{\infty} d\tau g_\eta(\tau)e^{-i\omega\tau}$, with the normalization $\int_{-\infty}^{\infty} d\omega S_\eta(\omega) = 2\pi$, and similarly for $S_s(\omega)$. We consider the background noise spectrum to be a Lorentzian, corresponding to $g_\eta(t) = e^{-|t|/\sigma_t}$, where σ_t denotes the correlation time-scale. The signal power-spectrum $S_s(\omega)$ is chosen to be white-cutoff, centered around a frequency Ω_0 and width $2\Delta\Omega$.

The aim of this protocol is to optimally detect the presence of a stochastic signal with a known spectrum in the presence of background noise. The detection protocol can be described in four steps. (i) Initialize the qubit in the state $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$. (ii) Let the qubit evolve for time $t = t_{\text{opt}}$, which we specify how to find later, in the presence of the Hamiltonian, $H(t)$. (iii) Rotate the qubit using a Hadamard gate U_{Hadamard}. (iv) Measure in the $\hat{\sigma}_z$ basis. Record outcome as '0' or '1'. Repeat these steps N_{shots} times.

Examining the dephasing dynamics in the presence of control, the probability to observe the qubit in state $|0\rangle$ is $P_{|0\rangle}(t) = \frac{1}{2}(1 + e^{-\chi(t)})$. Clearly, the dephasing exponent depends on the presence or absence of a signal and is denoted as $\chi_{\eta+s}$ and χ_η respectively. The corresponding outcome probabilities are labeled as $P_{\eta+s}$ and P_η respectively. In the following, we compute these two exponents in the regime of weak noise and signal as well as low SNR, and identify the control protocol that at a particular optimal time (t_{opt}) of measurement gives the maximum separation between the two decaying average probabilities, $\Delta P_{|0\rangle} = P_\eta - P_{\eta+s}$.

Both the signal and the noise cause the qubit to dephase, and the dephasing exponent is straightforwardly obtained in the Second Cumulant Approximation (SCA) [126]. Since the signal and noise are uncorrelated, the decay in the presence of a signal is a sum, $\chi_{\eta+s} = \chi_\eta + \chi_s$, where χ_s and χ_η are the decay exponents obtained from having just the signal or the noise present, respectively. Thus, the qubit decays at a faster rate in the presence of both signal and noise compared to just the background noise. In order to optimize for detection, we maximize over the difference between the two outcome probabilities, $\Delta P_{|0\rangle}$. We now have an effective

heuristic for designing optimal detection controls that becomes optimal when the SCA applies. We define the following objective function,

$$\mathcal{O}(t, \{\mathbf{\Omega}\}) = \Delta P_{|0\rangle} = \frac{1}{2} e^{-\chi_\eta(t)} \left(1 - e^{-\chi_s(t)}\right). \tag{A.2}$$

We carry out the optimization as a two-step procedure: (i) Optimize over control trajectories to obtain $\mathbf{\Omega}_{\text{opt}} = \text{argmax}_{\{\Omega\}} [\log \mathcal{O}]$ at a fixed detection time t. Note that the detection time sets the dimension of the control vector, $\dim [\mathbf{\Omega}] = t/\Delta t$ as we are considering discrete time steps. The optimization is implemented using stochastic gradient descent (SGD) algorithms. We do this using the Adam optimization algorithm implemented in TensorFlow library [100, 127]. Note that it is also straightforward to add additional constraints on the variables that could be motivated by experimental requirements. (ii) Optimize over the time of detection, t_{opt} to obtain the optimal detection protocol. The optimal time is obtained using a grid search over different detection times t. Therefore, we obtain the optimal detector in the SCA, which we denote as 'SCA-optimal', with $\mathcal{O}_{\text{opt}} = \mathcal{O}(t_{\text{opt}}, \mathbf{\Omega}_{\text{opt}})$.

Specifically, we do a grid search over different durations of the protocol, and for each t, we find the control schedule that optimizes the objective function. Since the cost function is differentiable, we use the Adam [100] optimizer with default parameters in TensorFlow to optimize the object function taking advantage of graphical processing units to accelerate the optimization procedure. Moreover, we constrain $[\mathbf{\Omega}]_i$ to be positive for all values of i. We either run the optimization for 10000 iterations, or stop if the magnitude of the difference of the objective function values

separated by 1000 iterations is smaller than 10^{-6}. In all the optimizations performed in this Appendix, we discretize the control to $\Delta t = 0.01J^{-1}$. We repeat this procedure for different control lengths t from $300\Delta t$ to $1300\Delta t$ in increments of $10\Delta t$. We then choose a t that has the optimal objective value.

After finishing this two-step optimization, we evaluate $\Delta P_{|0\rangle}$ at the optimal time with the optimal control by interpolating the results to a finer discretization of $\Delta t = 0.001J^{-1}$ and compare it to the value of $\Delta P_{|0\rangle}$ with other known control pulses such as CPMG and spin-lock. We obtain the best time for Carr, Purcell, Meiboom, Gill (CPMG) and spin-lock control schemes by numerically evaluating the $\Delta P_{|0\rangle}$ using discretized control with $\Delta t = 0.001J^{-1}$ and varying the duration of the protocol for t from $3000\Delta t$ to $13000\Delta t$ in increments of $100\Delta t$. The value of $\Delta P_{|0\rangle}$ corresponding to the numerically optimized control, CPMG, and spin-lock are compared in Fig. A.1.

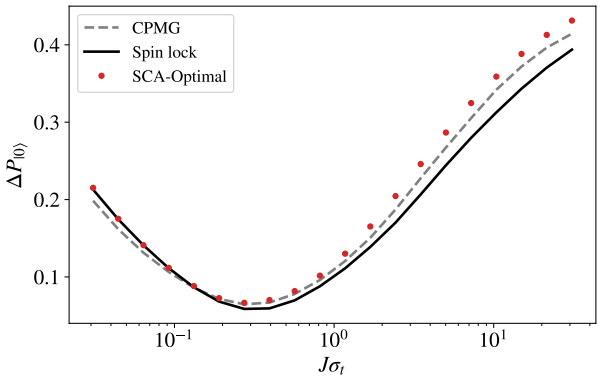

Figure A.1: Optimizing for detection for correlated background noise (Lorentzian spectrum) in the SCA. The signal is at $\Omega_0/J = 10\sqrt{\pi/30}$. Maximum of $\Delta P_{|0\rangle}$ as a function of background noise correlation, σ_t, for three different controls is shown.

The optimal protocol compared with Spin-lock and CPMG for different values of σ_t, together with their corresponding filter functions, are shown in Fig. A.2.

To take experimental limitations into account, we also consider a scenario where the maximum power is limited. This is manifested in a constraint on the maximum value of Ω at any time. We can implement this constraint by either adding an $L2$ regularization term to the objective function, that is $\mathcal{O}_{L2} = \beta||\mathbf{\Omega}||_2^2$, where β is a hyper parameter that controls the magnitude, or imposing a hard cutoff Ω_c on the value of the optimization variable. We observe that in a regime where the optimizer finds a CPMG-like control scheme to be optimal, adding these constraint transforms instantaneous pulses to finite-width version of the pulse that ultimately reaches the spin-lock solution, see Fig. A.3.

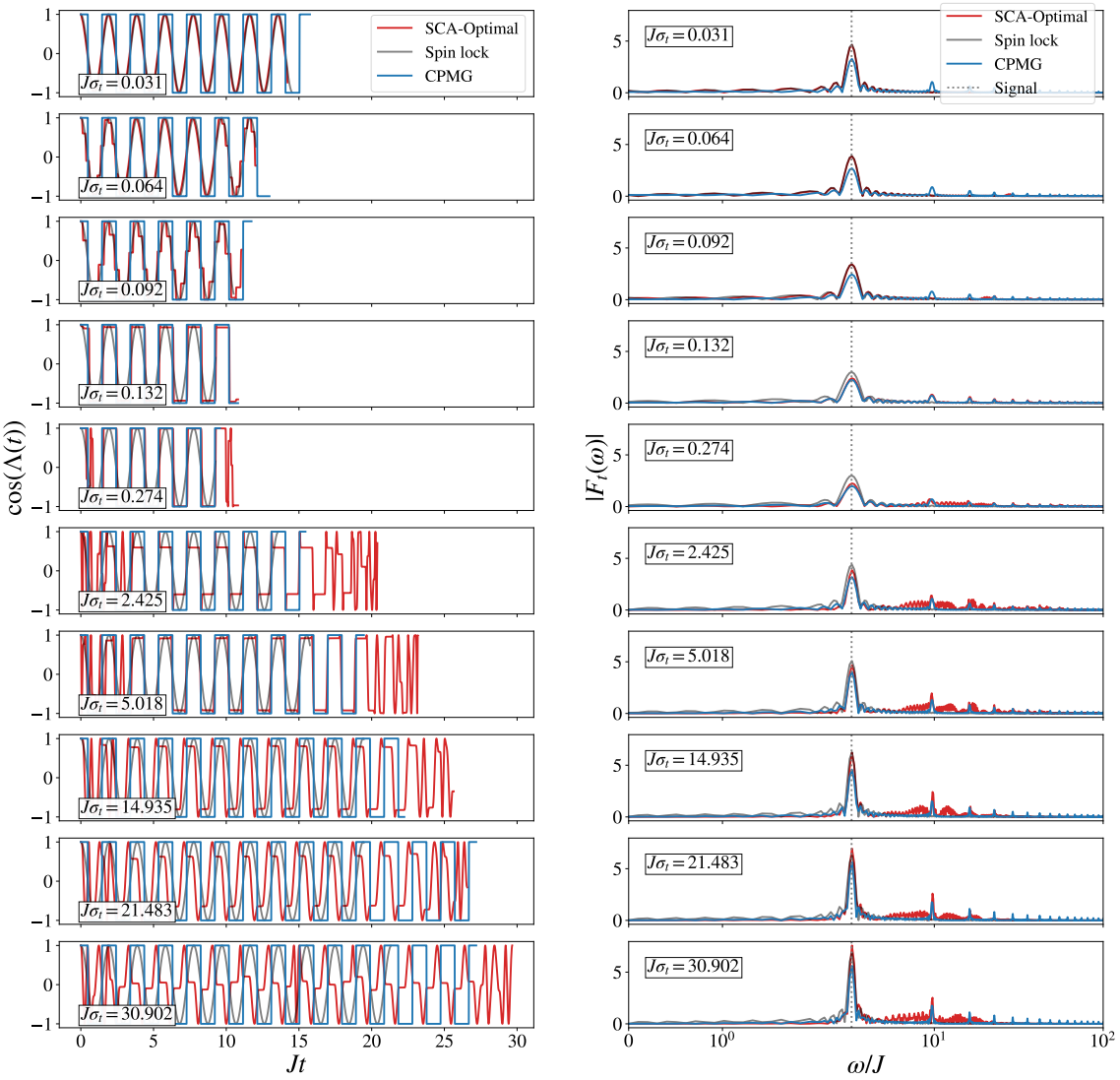

Figure A.2: Comparison of the numerically optimized protocol (SCA-optimal) with spin lock and CPMG protocols for different values of noise correlation σ_t. The left panel shows the time integrated control protocol $\cos(\Lambda(t))$, and the right panel shows the associated filter functions $|F_t(\omega)|$. Note that for small values of $J\sigma_t$, noise spectrum is nearly white and the optimal control coincides with spin-lock protocol. As $J\sigma_t$ increases the optimal control gets closer to CMPG protocol, and later assumes novel solutions that differ from both spin-lock and CPMG protocols.

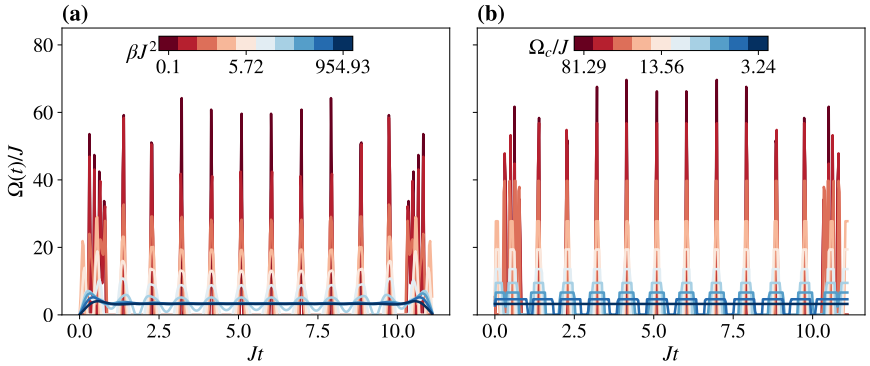

Figure A.3: The effect of constraints on the optimal protocol for $J\sigma_t = 0.56$. (a) With $L2$ regularization, we observe that as β is increased, the pulse solution gets smoother and pulses are widened. (b) A hard cutoff Ω_c is imposed on the amplitude of the drive. As the cutoff value is decreased the instantaneous pulses are transformed to finite width pulses.

Appendix B: Towards analog quantum simulations of lattice gauge theories with trapped ions

This chapter is based on the author's contribution to Ref. [128].

B.1 Summary of the results

Gauge field theories play a central role in modern physics and are at the heart of the Standard Model of elementary particles and interactions. Despite significant progress in applying classical computational techniques to simulate gauge theories, it has remained a challenging task to compute the real-time dynamics of systems described by gauge theories. An exciting possibility that has been explored in recent years is the use of highly controlled quantum systems to simulate, in an analog fashion, properties of a target system whose dynamics are difficult to compute. Engineered atom-laser interactions in a linear crystal of trapped ions offer a wide range of possibilities for quantum simulations of complex physical systems. Here we devise practical proposals for analog simulation of simple lattice gauge theories whose dynamics can be mapped onto spin-spin interactions in any dimension. These include $1+1D$ quantum electrodynamics, $2+1D$ Abelian Chern-Simons theory coupled to fermions, and $2+1D$ pure Z_2 gauge theory. The scheme proposed, along with the

optimization protocol applied, will have applications beyond the examples presented in this work, and will enable scalable analog quantum simulation of Heisenberg spin models in any number of dimensions and with arbitrary interaction strengths.

B.2 Tuning spin-dependent forces

Generally, the spin-dependent force in an trapped ion simulator leads to a spin-ion interactions in the form $(a_m + a_m^\dagger)(\alpha_0 I + \alpha_3 \sigma_z)$, where a_m is the annihilation operator corresponding to the mth phonon mode. To engineer some of the gauge theory Hamiltonians considered in Ref. [128], we have to engineer a spin dependent force that results in a spin-ion interactions in the form $(a_m + a_m^\dagger)\sigma_z$, where a_m is the annihilation operator corresponding to the mth phonon mode. Therefore, we have to find the laser polarization and amplitude that results in $\alpha_0 = 0$ and $\alpha_3 = 1$. Note that this condition is equivalent to having a spin dependent force such that $F_\uparrow = -F_\downarrow$.

The qubit is encoded in the magnetically-sensitive $|\uparrow\rangle \equiv |F = 0, m_F = 0\rangle$ and $|\downarrow\rangle \equiv |F = 1, m_F = -1\rangle$ hyperfine $^2S_{1/2}$ states of $^{171}\mathrm{Yb}^+$. Consider a set of Raman beams with frequencies ω_r and ω_b, detuned from $^2P_{1/2}$ manifold by Δ. In order to produce a spin-dependent force , the beams have to be detuned from each other by the motional mode's frequency ω_m, that is $\Delta\omega = \omega_b - \omega_r = \omega_m$, see Fig. B.1. In order to find appropriate polarizations and detuning that allow a pure σ_z Hamiltonian, three quantities must be calculated in this scheme: i) the Stark shift induced by red and blue lasers in the Raman pair, ii) the spontaneous emission

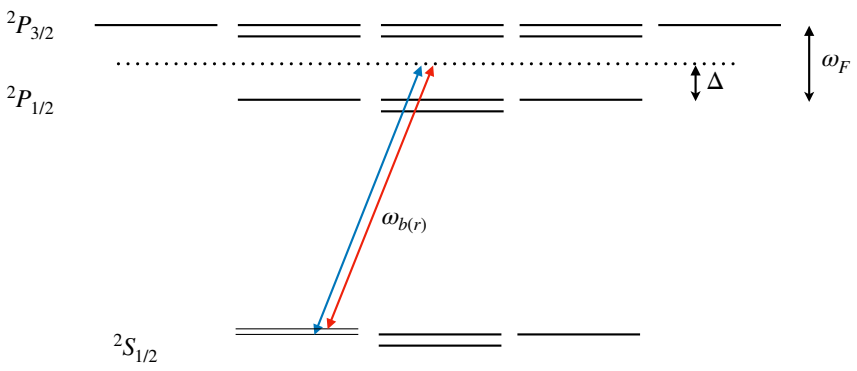

Figure B.1: The level diagram of $^{171}Yb^+$ relevant to the scheme presented in this Appendix.

rate from excited states, and finally iii) the spin-dependent force on the qubit. (iii) must be studied to deduce the conditions under which $F_\uparrow = -F_\downarrow$, while at the same time (i) must be ensured to vanish, and (ii) must be minimized.

Let us denote the polarization of each beam by $\hat{\boldsymbol{\epsilon}}_r = r_-\hat{\sigma}_- + r_0\hat{\pi} + r_+\hat{\sigma}_+$ and $\hat{\boldsymbol{\epsilon}}_b = b_-\hat{\sigma}_- + b_0\hat{\pi} + b_+\hat{\sigma}_+$, where $|r_-|^2 + |r_0|^2 + |r_+|^2 = |b_-|^2 + |b_0|^2 + |b_+|^2 = 1$. In calculating these quantities, matrix elements in the form $\langle \alpha' F' m'_F | \boldsymbol{d} \cdot \hat{\boldsymbol{\epsilon}} | \alpha F m_F \rangle$ need to be evaluated, where $\boldsymbol{d}$ is the electric dipole operator, and α represents all other quantum numbers of the state besides the total spin F (nuclear spin added to electron's total angular momentum) and its component along the quantization axis, m_F. Such a matrix element can be evaluated using [129]

$$\langle \alpha' F' m'_F | \boldsymbol{d}.\hat{\boldsymbol{\epsilon}} | \alpha F m_F \rangle = (-1)^{J'+I-m'_F} \sqrt{(2F+1)(2F'+1)} \begin{Bmatrix} J' & F' & I \\ F & J & 1 \end{Bmatrix} \begin{pmatrix} F & 1 & F' \\ m_F & q & -m'_F \end{pmatrix} \langle \alpha' J' || d || \alpha J \rangle. \tag{B.1}$$

Here, $q = -1$ for the $\hat{\sigma}_-$-polarized light, $q = 0$ for the $\hat{\pi}$-polarized light, and $q = 1$ for the $\hat{\sigma}_+$-polarized light. I and J denote the total nuclear spin and the electron's total angular momentum, respectively. "()" corresponds Wigner's $3j$ symbol while "{}" corresponds to Wigner's $6j$ symbols. The reduced matrix element $\langle \alpha' J' || d || \alpha J \rangle$ is related to the spontaneous emission rate γ between states with J and J' quantum numbers for an atom coupled to free space:

$$|\langle \alpha' J' || d || \alpha J \rangle|^2 = c_0 (2J' + 1)\gamma, \tag{B.2}$$

where c_0 is a number that depends on the transitions. For simplicity, in the following we assume that the ${}^2P_{1/2}$ and the ${}^2P_{3/2}$ states have the same c_0 and γ.

B.2.1 Stark shift

In the limit where $\Delta \gg \gamma$, the Stark shift for $|m_S\rangle = |\uparrow\rangle, |\downarrow\rangle$ is given by [130]

$$\delta_{\text{Stark}}(m_S) = \frac{1}{4} \sum_{j=r,b} \sum_i \frac{|\langle m_S | \boldsymbol{d} \cdot \hat{\boldsymbol{\epsilon}}_j E_j | i \rangle|^2}{\Delta_i}, \tag{B.3}$$

where Δ_i is the detuning from the states that are virtually occupied, and E_j is the electric-field amplitude. Using Eq. (B.1), the net Stark shift is found to be

$$\begin{aligned} \delta_{\text{Stark}}(\uparrow) - \delta_{\text{Stark}}(\downarrow) = & \frac{c_0 \gamma \omega_F}{12^2 \Delta(\Delta - \omega_F)} \\ & \times \left(|b_-|^2 + |r_-|^2 - |b_+|^2 - |r_+|^2 \right). \end{aligned} \tag{B.4}$$

As is evident, by choosing $|b_-|^2 + |r_-|^2 = |b_+|^2 + |r_+|^2$, the net shift can be set to zero.

B.2.2 Spontaneous emission

The spontaneous emission rate can be evaluated using [130]

$$R_{\mathrm{SE}} = \frac{1}{4}\sum_i \sum_{j=r,b} \sum_{m_S=\downarrow,\uparrow} \frac{P_{m_S}\gamma_i |\langle m_S|\boldsymbol{d}\cdot\hat{\boldsymbol{\epsilon}}_j E_j|i\rangle|^2}{\Delta_i^2}, \tag{B.5}$$

where P_{m_S} is the probability of being in the m_S ground state. Under the constraint that sets Eq. (B.4) to zero, one finds that

$$R_{\mathrm{SE}} = \frac{c_0\gamma^2(2+|r_0|^2+|b_0|^2)}{12\sqrt{(1+|r_0|^2)(1+|b_0|^2)}}\left[\frac{1}{\Delta^2}+\frac{2}{(\Delta-\omega_F)^2}\right]. \tag{B.6}$$

As is seen, with the choice $\Delta = (\sqrt{2}-1)\omega_F$ one is close to a local minimum of the spontaneous emission rate.

B.2.3 Spin-dependent force

Finally, the spin-dependent force can be found by considering the resonant two-photon Raman Rabi rate [130]

$$\Omega(m_S) = \frac{e^{i(\varphi_b-\varphi_r)}}{4}\sum_i \frac{\langle m_S|\boldsymbol{d}\cdot\hat{\boldsymbol{\epsilon}}_r E_r|i\rangle\langle i|\boldsymbol{d}\cdot\hat{\boldsymbol{\epsilon}}_b E_b|m_S\rangle}{\Delta_i}, \tag{B.7}$$

where φ_r and φ_b are the phases of the red- and blue-detuned beams, respectively. With $\Delta\varphi \equiv \varphi_b - \varphi_r = 0$ and $\Delta = (\sqrt{2}-1)\omega_F$, one find that

$$\Omega(\downarrow) = \frac{-\gamma(b_0 r_0^* + b_- r_-^* + b_+ r_+^*)}{12\omega_F}, \tag{B.8}$$

$$\Omega(\uparrow) = \frac{\gamma(-2b_0 r_0^* + (2+3\sqrt{2})b_+ r_+^* - 3(2+\sqrt{2})b_- r_-^*)}{24\omega_F}. \tag{B.9}$$

In order to satisfy the condition $\Omega(\downarrow) = -\Omega(\uparrow)$ or in turn $F_\uparrow = -F_\downarrow$, since the spin-dependent force is related to the Rabi frequency via $F_{m_S} = \Delta k\ \Omega(m_S)$. A choice for the polarization vectors is

$$\hat{\boldsymbol{\epsilon}}_b = \frac{3}{2-\sqrt{2}}(-1, \sqrt{2+\frac{3}{\sqrt{2}}}, 1), \tag{B.10}$$

$$\hat{\boldsymbol{\epsilon}}_r = \frac{3}{2-\sqrt{2}}(1, \sqrt{2+\frac{3}{\sqrt{2}}}, 1). \tag{B.11}$$

Of course, these analytical solutions rely on the approximations that were made throughout these calculations, such as equal spontaneous emission rate from all the excited states considered. When precise values of the physical parameters in the system are input, the optimal values for the parameters can still be evaluated numerically using the formalism outlined.

Bibliography

[1] Chris Monroe, DM Meekhof, BE King, Wayne M Itano, and David J Wineland. Demonstration of a fundamental quantum logic gate. *Physical review letters*, 75(25):4714, 1995.

[2] Shantanu Debnath, Norbert M Linke, Caroline Figgatt, Kevin A Landsman, Kevin Wright, and Christopher Monroe. Demonstration of a small programmable quantum computer with atomic qubits. *Nature*, 536(7614):63–66, 2016.

[3] Ryszard Horodecki, Paweł Horodecki, Michał Horodecki, and Karol Horodecki. Quantum entanglement. *Rev. Mod. Phys.*, 81:865–942, Jun 2009.

[4] P. W. Shor. Algorithms for quantum computation: discrete logarithms and factoring. In *Proceedings 35th Annual Symposium on Foundations of Computer Science*, pages 124–134, 1994.

[5] C. L. Degen, F. Reinhard, and P. Cappellaro. Quantum sensing. *Rev. Mod. Phys.*, 89:035002, Jul 2017.

[6] J. J. Sakurai and Jim Napolitano. *Modern Quantum Mechanics*. Cambridge University Press, 2 edition, 2017.

[7] Heinz-Peter Breuer, Francesco Petruccione, et al. *The theory of open quantum systems*. Oxford University Press on Demand, 2002.

[8] Vittorio Gorini, Andrzej Kossakowski, and Ennackal Chandy George Sudarshan. Completely positive dynamical semigroups of n-level systems. *Journal of Mathematical Physics*, 17(5):821–825, 1976.

[9] Goran Lindblad. On the generators of quantum dynamical semigroups. *Communications in Mathematical Physics*, 48(2):119–130, 1976.

[10] Norman F. Ramsey. A molecular beam resonance method with separated oscillating fields. *Phys. Rev.*, 78:695–699, Jun 1950.

[11] Kaveh Khodjasteh, Daniel A Lidar, and Lorenza Viola. Arbitrarily accurate dynamical control in open quantum systems. *Physical review letters*, 104(9):090501, 2010.

[12] Lorenza Viola, Emanuel Knill, and Seth Lloyd. Dynamical decoupling of open quantum systems. *Physical Review Letters*, 82(12):2417, 1999.

[13] A Bermudez, Piet O Schmidt, Martin B Plenio, and A Retzker. Robust trapped-ion quantum logic gates by continuous dynamical decoupling. *Physical Review A*, 85(4):040302, 2012.

[14] Sebastian Diehl, A Micheli, A Kantian, B Kraus, HP Büchler, and P Zoller. Quantum states and phases in driven open quantum systems with cold atoms. *Nature Physics*, 4(11):878–883, 2008.

[15] Tomaž Prosen and Iztok Pižorn. Quantum phase transition in a far-from-equilibrium steady state of an x y spin chain. *Physical review letters*, 101(10):105701, 2008.

[16] Sebastian Diehl, Andrea Tomadin, Andrea Micheli, Rosario Fazio, and Peter Zoller. Dynamical phase transitions and instabilities in open atomic many-body systems. *Phys. Rev. Lett.*, 105:015702, Jul 2010.

[17] Christopher Jarzynski. Equalities and inequalities: Irreversibility and the second law of thermodynamics at the nanoscale. *Annu. Rev. Condens. Matter Phys.*, 2(1):329–351, 2011.

[18] Gavin E Crooks. Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Physical Review E*, 60(3):2721, 1999.

[19] Udo Seifert. Stochastic thermodynamics, fluctuation theorems and molecular machines. *Reports on Progress in Physics*, 75(12):126001, 2012.

[20] Edward H Feng and Gavin E Crooks. Length of time's arrow. *Physical review letters*, 101(9):090602, 2008.

[21] Andrea Hofmann, Ville F Maisi, Julien Basset, Christian Reichl, Werner Wegscheider, Thomas Ihn, Klaus Ensslin, and Christopher Jarzynski. Heat dissipation and fluctuations in a driven quantum dot. *physica status solidi (b)*, 254(3):1600546, 2017.

[22] Jens Eisert, Dominik Hangleiter, Nathan Walk, Ingo Roth, Damian Markham, Rhea Parekh, Ulysse Chabaud, and Elham Kashefi. Quantum certification and benchmarking. *Nature Reviews Physics*, 2(7):382–390, Jul 2020.

[23] Giuseppe Carleo, Ignacio Cirac, Kyle Cranmer, Laurent Daudet, Maria Schuld, Naftali Tishby, Leslie Vogt-Maranto, and Lenka Zdeborová. Machine learning and the physical sciences. *Reviews of Modern Physics*, 91(4):045002, 2019.

[24] Jessica Montgomery. The ai revolution in science: applications and new research directions, Aug 2019.

[25] Paraj Titum, Kevin M Schultz, Alireza Seif, Gregory D Quiroz, and BD Clader. Optimal control for quantum detectors. *arXiv preprint arXiv:2005.05995*, 2020.

[26] Alireza Seif, Kevin A Landsman, Norbert M Linke, Caroline Figgatt, C Monroe, and Mohammad Hafezi. Machine learning assisted readout of trapped-ion qubits. *Journal of Physics B: Atomic, Molecular and Optical Physics*, 51(17):174006, 2018.

[27] Zohreh Davoudi, Mohammad Hafezi, Christopher Monroe, Guido Pagano, Alireza Seif, and Andrew Shaw. Towards analog quantum simulations of lattice gauge theories with trapped ions. *Phys. Rev. Research*, 2:023015, Apr 2020.

[28] Alireza Seif, Mohammad Hafezi, and Christopher Jarzynski. Machine learning the thermodynamic arrow of time. *Nature Physics*, Sep 2020.

[29] Mark A. Davenport, Marco F. Duarte, Yonina C. Eldar, and Gitta Kutyniok. Introduction to compressed sensing. In Yonina C. Eldar and Gitta Kutyniok, editors, *Compressed Sensing: Theory and Applications*, pages 1–64. Cambridge University Press, 2012.

[30] Emmanuel Candès. The restricted isometry property and its implications for compressed sensing. *Compte Rendus de l'Academie des Sciences*, 346:589–592, 05 2008.

[31] Roman Vershynin. Introduction to the non-asymptotic analysis of random matrices. In Yonina C. Eldar and Gitta Kutyniok, editors, *Compressed Sensing: Theory and Applications*, page 210–268. Cambridge University Press, 2012.

[32] Holger Rauhut. Compressive sensing and structured random matrices. *Theoretical foundations and numerical methods for sparse recovery*, 9:1–92, 2010.

[33] Michel Ledoux and Michel Talagrand. *Probability in Banach Spaces: isoperimetry and processes*. Springer Science & Business Media, 2013.

[34] Mark Rudelson and Roman Vershynin. On sparse reconstruction from fourier and gaussian measurements. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 61(8):1025–1045, 2008.

[35] Yi-Kai Liu. Universal low-rank matrix recovery from pauli measurements. In *Advances in Neural Information Processing Systems*, pages 1638–1646, 2011.

[36] Steven Diamond and Stephen Boyd. CVXPY: A Python-embedded modeling language for convex optimization. *Journal of Machine Learning Research*, 17(83):1–5, 2016.

[37] Akshay Agrawal, Robin Verschueren, Steven Diamond, and Stephen Boyd. A rewriting system for convex optimization problems. *Journal of Control and Decision*, 5(1):42–60, 2018.

[38] Bernhard Baumgartner. An inequality for the trace of matrix products, using absolute values. *arXiv preprint arXiv:1106.6189*, 2011.

[39] Fuzhen Zhang. *Matrix theory: basic results and techniques*. Springer Science & Business Media, 2011.

[40] Gerardo A Paz-Silva, Leigh M Norris, and Lorenza Viola. Multiqubit spectroscopy of gaussian quantum noise. *Physical Review A*, 95(2):022121, 2017.

[41] Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone. Quantum metrology. *Physical review letters*, 96(1):010401, 2006.

[42] Peter W. Shor. Scheme for reducing decoherence in quantum computer memory. *Phys. Rev. A*, 52:R2493–R2496, Oct 1995.

[43] Andrew Steane. Multiple-particle interference and quantum error correction. *Proc. R. Soc. Lond. A*, 452(1954):2551–2577, 1996.

[44] Daniel Gottesman. An introduction to quantum error correction and fault-tolerant quantum computation. *arXiv preprint arXiv:0904.2557*, 2009.

[45] AH Myerson, DJ Szwer, SC Webster, DTC Allcock, MJ Curtis, G Imreh, JA Sherman, DN Stacey, AM Steane, and DM Lucas. High-fidelity readout of trapped-ion qubits. *Physical Review Letters*, 100(20):200502, 2008.

[46] Evan Jeffrey, Daniel Sank, JY Mutus, TC White, J Kelly, R Barends, Y Chen, Z Chen, B Chiaro, A Dunsworth, et al. Fast accurate state measurement with superconducting qubits. *Physical review letters*, 112(19):190504, 2014.

[47] S Olmschenk, KC Younge, DL Moehring, DN Matsukevich, P Maunz, and C Monroe. Manipulation and detection of a trapped yb+ hyperfine qubit. *Physical Review A*, 76(5):052314, 2007.

[48] Alice Heather Burrell. *High fidelity readout of trapped ion qubits*. PhD thesis, Oxford University, UK.

[49] Michael I Jordan and Tom M Mitchell. Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245):255–260, 2015.

[50] Michael A Nielsen. Neural networks and deep learning, 2015.

[51] Giuseppe Carleo and Matthias Troyer. Solving the quantum many-body problem with artificial neural networks. *Science*, 355(6325):602–606, 2017.

[52] Hiroki Saito. Solving the bose–hubbard model with machine learning. *Journal of the Physical Society of Japan*, 86(9):093001, 2017.

[53] Xun Gao and Lu-Ming Duan. Efficient representation of quantum many-body states with deep neural networks. *Nature communications*, 8(1):662, 2017.

[54] Dong-Ling Deng, Xiaopeng Li, and S Das Sarma. Quantum entanglement in neural network states. *Physical Review X*, 7(2):021021, 2017.

[55] Dong-Ling Deng, Xiaopeng Li, and S. Das Sarma. Machine learning topological states. *Phys. Rev. B*, 96:195145, Nov 2017.

[56] Giacomo Torlai, Guglielmo Mazzola, Juan Carrasquilla, Matthias Troyer, Roger Melko, and Giuseppe Carleo. Neural-network quantum state tomography. *Nature Physics*, page 1, 2018.

[57] Paul Baireuther, Thomas E O'Brien, Brian Tarasinski, and Carlo WJ Beenakker. Machine-learning-assisted correction of correlated qubit errors in a topological code. *Quantum*, 2:48, 2018.

[58] Stefan Krastanov and Liang Jiang. Deep neural network probabilistic decoder for stabilizer codes. *Scientific reports*, 7(1):11003, 2017.

[59] Giacomo Torlai and Roger G Melko. Neural decoder for topological codes. *Physical Review Letters*, 119(3):030501, 2017.

[60] Nishad Maskara, Aleksander Kubica, and Tomas Jochym-O'Connor. Advantages of versatile neural-network decoding for topological codes. *arXiv preprint arXiv:1802.08680*, 2018.

[61] Evert PL van Nieuwenburg, Ye-Hua Liu, and Sebastian D Huber. Learning phase transitions by confusion. *Nature Physics*, 13(5):435, 2017.

[62] Evert van Nieuwenburg, Eyal Bairey, and Gil Refael. Learning phase transitions from dynamics. *arXiv preprint arXiv:1712.00450*, 2017.

[63] Juan Carrasquilla and Roger G Melko. Machine learning phases of matter. *Nature Physics*, 13(5):431–434, 2017.

[64] Ian T Jolliffe and Jorge Cadima. Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065):20150202, 2016.

[65] Rui Xu and Donald Wunsch. Survey of clustering algorithms. *IEEE Transactions on neural networks*, 16(3):645–678, 2005.

[66] Sebastian Johann Wetzel. Unsupervised learning of phase transitions: from principle component analysis to variational autoencoders. *arXiv preprint arXiv:1703.02435*, 2017.

[67] Easwar Magesan, Jay M Gambetta, AD Córcoles, and Jerry M Chow. Machine learning for discriminating quantum measurement trajectories and improving readout. *Physical review letters*, 114(20):200501, 2015.

[68] M Acton, K-A Brickman, PC Haljan, PJ Lee, L Deslauriers, and C Monroe. Near-perfect simultaneous measurement of a qubit register. *Quantum Information & Computation*, 6(6):465–482, 2006.

[69] Stephen Gregory Crain. *Integrated System Technologies for Modular Trapped Ion Quantum Information Processing*. PhD thesis, Duke University, USA, 2016.

[70] Chao Shen and LM Duan. Correcting detection errors in quantum state engineering through data processing. *New Journal of Physics*, 14(5):053053, 2012.

[71] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436, 2015.

[72] Matthew D Zeiler. Adadelta: an adaptive learning rate method. *arXiv preprint arXiv:1212.5701*, 2012.

[73] Felix A Gers, Nicol N Schraudolph, and Jürgen Schmidhuber. Learning precise timing with lstm recurrent networks. *Journal of machine learning research*, 3(Aug):115–143, 2002.

[74] Arthur Stanley Eddington. *The nature of the physical world,*. New York,The Macmillan Company, 1928.

[75] Édgar Roldán, Izaak Neri, Meik Dörpinghaus, Heinrich Meyr, and Frank Jülicher. Decision making in the arrow of time. *Phys. Rev. Lett.*, 115:250602, Dec 2015.

[76] Gavin E. Crooks. Nonequilibrium measurements of free energy differences for microscopically reversible markovian systems. *Journal of Statistical Physics*, 90(5):1481–1487, Mar 1998.

[77] Andrew W. Senior, Richard Evans, John Jumper, James Kirkpatrick, Laurent Sifre, Tim Green, Chongli Qin, Augustin Žídek, Alexander W. R. Nelson, Alex Bridgland, Hugo Penedones, Stig Petersen, Karen Simonyan, Steve Crossan, Pushmeet Kohli, David T. Jones, David Silver, Koray Kavukcuoglu, and Demis Hassabis. Improved protein structure prediction using potentials from deep learning. *Nature*, 577(7792):706–710, 2020.

[78] Giacomo Torlai and Roger G. Melko. Learning thermodynamics with boltzmann machines. *Phys. Rev. B*, 94:165134, Oct 2016.

[79] Sebastian J. Wetzel and Manuel Scherzer. Machine learning of explicit order parameters: From the ising model to su(2) lattice gauge theory. *Phys. Rev. B*, 96:184410, Nov 2017.

[80] Kelvin Ch'ng, Juan Carrasquilla, Roger G. Melko, and Ehsan Khatami. Machine learning phases of strongly correlated fermions. *Phys. Rev. X*, 7:031038, Aug 2017.

[81] Kelvin Ch'ng, Nick Vazquez, and Ehsan Khatami. Unsupervised machine learning account of magnetic transitions in the hubbard model. *Phys. Rev. E*, 97:013306, Jan 2018.

[82] Ye-Hua Liu and Evert PL van Nieuwenburg. Discriminative cooperative networks for detecting phase transitions. *Physical review letters*, 120(17):176401, 2018.

[83] Frank Schindler, Nicolas Regnault, and Titus Neupert. Probing many-body localization with neural networks. *Physical Review B*, 95(24):245134, 2017.

[84] Louis-François Arsenault, Alejandro Lopez-Bezanilla, O Anatole von Lilienfeld, and Andrew J Millis. Machine learning for many-body physics: the case of the anderson impurity model. *Physical Review B*, 90(15):155136, 2014.

[85] Matthew JS Beach, Anna Golubeva, and Roger G Melko. Machine learning vortices at the kosterlitz-thouless transition. *Physical Review B*, 97(4):045207, 2018.

[86] Evert van Nieuwenburg, Eyal Bairey, and Gil Refael. Learning phase transitions from dynamics. *Phys. Rev. B*, 98:060301, Aug 2018.

[87] Pedro Ponte and Roger G. Melko. Kernel methods for interpretable machine learning of order parameters. *Phys. Rev. B*, 96:205146, Nov 2017.

[88] Michael Schmidt and Hod Lipson. Distilling free-form natural laws from experimental data. *Science*, 324(5923):81–85, 2009.

[89] Samuel H Rudy, Steven L Brunton, Joshua L Proctor, and J Nathan Kutz. Data-driven discovery of partial differential equations. *Science Advances*, 3(4):e1602614, 2017.

[90] Raban Iten, Tony Metger, Henrik Wilming, Lídia Del Rio, and Renato Renner. Discovering physical concepts with neural networks. *Physical Review Letters*, 124(1):010508, 2020.

[91] Joseph Berkson. Application of the logistic function to bio-assay. *Journal of the American statistical association*, 39(227):357–365, 1944.

[92] N. E. Day and D. F. Kerridge. A general maximum likelihood discriminant. *Biometrics*, 23(2):313–323, 1967.

[93] Christopher Jarzynski. Nonequilibrium equality for free energy differences. *Physical Review Letters*, 78(14):2690, 1997.

[94] Christopher Jarzynski. Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach. *Physical Review E*, 56(5):5018, 1997.

[95] Gerhard Hummer and Attila Szabo. Free energy reconstruction from nonequilibrium single-molecule pulling experiments. *Proceedings of the National Academy of Sciences (USA)*, 98(7):3658–3661, 2001.

[96] Michael R. Shirts, Eric Bair, Giles Hooker, and Vijay S. Pande. Equilibrium free energies from nonequilibrium measurements using maximum-likelihood methods. *Phys. Rev. Lett.*, 91:140601, Oct 2003.

[97] Paul Maragakis, Felix Ritort, Carlos Bustamante, Martin Karplus, and Gavin E Crooks. Bayesian estimates of free energies from nonequilibrium work data in the presence of instrument noise. *The Journal of chemical physics*, 129(2):07B609, 2008.

[98] Christopher Jarzynski. Rare events and the convergence of exponentially averaged work values. *Phys. Rev. E*, 73:046105, Apr 2006.

[99] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

[100] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[101] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1):1929–1958, 2014.

[102] Rich Caruana. Multitask learning. *Machine learning*, 28(1):41–75, 1997.

[103] Alexander Mordvintsev, Christopher Olah, and Mike Tyka. Inceptionism: Going deeper into neural networks, 2015.

[104] Frank Schindler, Nicolas Regnault, and Titus Neupert. Probing many-body localization with neural networks. *Phys. Rev. B*, 95:245134, Jun 2017.

[105] Steven J. Nowlan and Geoffrey E. Hinton. Evaluation of adaptive mixtures of competing experts. In *NIPS*, 1990.

[106] Silvio R. A. Salinas. *Introduction to Statistical Physics*. Springer New York, New York, NY, 2001.

[107] Christopher W Wächtler, Philipp Strasberg, Sabine HL Klapp, Gernot Schaller, and Christopher Jarzynski. Stochastic thermodynamics of self-oscillations: the electron shuttle. *New Journal of Physics*, 21(7):073009, 2019.

[108] Jeremy T Young, Alexey V Gorshkov, Michael Foss-Feig, and Mohammad F Maghrebi. Nonequilibrium fixed points of coupled ising models. *Physical Review X*, 10(1):011039, 2020.

[109] Lyndsey C Pickup, Zheng Pan, Donglai Wei, YiChang Shih, Changshui Zhang, Andrew Zisserman, Bernhard Scholkopf, and William T Freeman. Seeing the arrow of time. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2035–2042, 2014.

[110] Donglai Wei, Joseph J Lim, Andrew Zisserman, and William T Freeman. Learning and using the arrow of time. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8052–8060, 2018.

[111] Joel J Wallman and Steven T Flammia. Randomized benchmarking with confidence. *New Journal of Physics*, 16(10):103032, 2014.

[112] Genyue Liu, Mo Chen, Yi-Xiang Liu, David Layden, and Paola Cappellaro. Repetitive readout enhanced by machine learning. *Machine Learning: Science and Technology*, 1(1):015003, 2020.

[113] Giacomo Torlai, Brian Timar, Evert PL van Nieuwenburg, Harry Levine, Ahmed Omran, Alexander Keesling, Hannes Bernien, Markus Greiner, Vladan Vuletić, Mikhail D Lukin, et al. Integrating neural networks with a quantum simulator for state reconstruction. *Physical review letters*, 123(23):230504, 2019.

[114] Sandesh S Kalantre, Justyna P Zwolak, Stephen Ragole, Xingyao Wu, Neil M Zimmerman, Michael D Stewart, and Jacob M Taylor. Machine learning techniques for state recognition and auto-tuning in quantum dots. *npj Quantum Information*, 5(1):1–10, 2019.

[115] H Moon, DT Lennon, J Kirkpatrick, NM van Esbroeck, LC Camenzind, Liuqi Yu, F Vigneau, DM Zumbühl, G Andrew D Briggs, MA Osborne, et al. Machine learning enables completely automatic tuning of a quantum device faster than human experts. *Nature communications*, 11(1):1–10, 2020.

[116] Jianwei Wang, Stefano Paesani, Raffaele Santagati, Sebastian Knauer, Antonio A Gentile, Nathan Wiebe, Maurangelo Petruzzella, Jeremy L O'Brien, John G Rarity, Anthony Laing, et al. Experimental quantum hamiltonian learning. *Nature Physics*, 13(6):551–555, 2017.

[117] Xuechen Li, Ting-Kam Leonard Wong, Ricky TQ Chen, and David Duvenaud. Scalable gradients for stochastic differential equations. *arXiv preprint arXiv:2001.01328*, 2020.

[118] Andrew Y Ng and Stuart J Russell. Algorithms for inverse reinforcement learning. In *ICML*, 2000.

[119] A.V. Getling. *Rayleigh-Benard Convection: Structures and Dynamics*. Advanced series in nonlinear dynamics. World Scientific, 1998.

[120] Aditi Mitra, So Takei, Yong Baek Kim, and AJ Millis. Nonequilibrium quantum criticality in open electronic systems. *Physical review letters*, 97(23):236808, 2006.

[121] Georgios Roumpos, Michael Lohse, Wolfgang H Nitsche, Jonathan Keeling, Marzena Hanna Szymańska, Peter B Littlewood, Andreas Löffler, Sven Höfling, Lukas Worschech, Alfred Forchel, et al. Power-law decay of the spatial correlation function in exciton-polariton condensates. *Proceedings of the National Academy of Sciences*, 109(17):6467–6472, 2012.

[122] Jeremy T. Young, Alexey V. Gorshkov, Michael Foss-Feig, and Mohammad F. Maghrebi. Nonequilibrium fixed points of coupled ising models. *Phys. Rev. X*, 10:011039, Feb 2020.

[123] Jamir Marino and Sebastian Diehl. Driven markovian quantum criticality. *Physical review letters*, 116(7):070407, 2016.

[124] Emanuele G Dalla Torre, Eugene Demler, Thierry Giamarchi, and Ehud Altman. Quantum critical states and phase transitions in the presence of non-equilibrium noise. *Nature Physics*, 6(10):806–810, 2010.

[125] Iris Cong, Soonwon Choi, and Mikhail D Lukin. Quantum convolutional neural networks. *Nature Physics*, 15(12):1273–1278, 2019.

[126] Paraj Titum, Kevin M Schultz, Alireza Seif, Gregory D Quiroz, and BD Clader. Optimal control for quantum detectors. *arXiv preprint arXiv:2005.05995*, 2020.

[127] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*, 2016.

[128] Zohreh Davoudi, Mohammad Hafezi, Christopher Monroe, Guido Pagano, Alireza Seif, and Andrew Shaw. Towards analog quantum simulations of lattice gauge theories with trapped ions. *Phys. Rev. Research*, 2:023015, Apr 2020.

[129] Jonathan Albert Mizrahi. *ULTRAFAST CONTROL OF SPIN AND MOTION IN TRAPPED IONS*. PhD thesis, University of Maryland, College Park, USA, 2013.

[130] David J Wineland, M Barrett, J Britton, J Chiaverini, B DeMarco, Wayne M Itano, B Jelenković, Christopher Langer, D Leibfried, V Meyer, et al. Quantum information processing with trapped ions. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 361(1808):1349–1361, 2003.