Stochastic and Neural Network methods for many-body open quantum systems

Filippo Vicentini

Université de Paris

Laboratoire Matériaux et Phénomènes Quantiques

École Doctorale 564 : Physique en Ile-de-France

Thèse de doctorat de Physique

Stochastic and Neural Network methods for many-body open quantum systems

presentée par

Filippo Vicentini

sous la direction de:

Prof. Cristiano Ciuti

Jury

PR	Sara DUCCI	Université de Paris	Prèsident
PR	Dieter JAKSCH	University of Oxford	Rapporteur
PR	Guido PUPILLO	Université de Strasbourg	Rapporteur
DR	Markus HOLTZMANN	Université de Grenoble Alpes	Examinateur
DR	Mazyar MIRRAHIMI	Inria Paris	Examinateur
PR	Cristiano CIUTI	Université de Paris	Directeur de thèse

Thèse défendue publiquement à Paris le 11 decembre 2019

A Marco e Marina

Acknowledgements

With the defence of this Ph.D. thesis I am also concluding another dynamic part of my life, characterised by an eclectic mixture of enthusiasm and depression, love and mental health crises, a lot of work, passion and a huge personal development. The overall successful tale of this doctorate owes several debts to the interwined web of people that surrounded. After all, the real outcome of this Doctorate in Physics is not the little knowledge contained in those pages, but the emergent beauty of the chaotic patchwork of connections and friendships that emerged. So much, in fact, that at times I believe I now hold Doctorate in Tailoring, having perfected the art of knitting together the personalities of the different characters I happened to find in that endless source that Paris is.

Within this colorful mosaic, I particularly love the melody created by the antiresonance of Cristiano's unfathomable optimism with my limitless critical pessimism. However, the only reason this worked out is because of the ton of hard work invested by those like Riccardo and, especially, Alberto, who devoted time to patiently walk me throughout all the chapters of 'how to be a physicist, for dummies'.

Of course, the real emotional load of the unlimited mental stress caused by the unhealthy academic societal structure lied on the shoulders of those who spent the rest of the working days with me, such as Titta and Silvia together with the squatters Francesca and Alberto, without whom I would probably not resisted so long. With them, I also think about those other people I owe lots to, such as the Thellinos, Lucia, Enrico and Rocco who (approximately) met one night in a night train between Paris and Milano, and the dancers, Marie-Laure and Claire, as well as Lucille, who so patiently taught me how to speak French and enjoy the parisian life.

Finally, this story could never have been lived if my family had not supported me through this endeavour that lasted more than 8 years, allowing me to focus on the beautiful questions that Physics and being young throw at you, while only stressing me a (not-really-so-)little bit about everything else that life threw at them and at me.

The moments that I shared with all the people I mentioned above, as well as the many unnamed ones, are the reason why I can confidently say that I would forever do this again and I'm grateful I decided to stay in Paris, a few years ago.

Thank you all,

Filippo

Summary

This thesis is devoted to the theoretical study of driven-dissipative many-body quantum systems with a particular focus on advanced computational methods. After reviewing several state-of-the-art numerical methods to determine the steady-state and dynamics of open quantum systems described by a Lindblad master equation for the density matrix, this thesis manuscript presents two original methods: (i) an optimal unraveling of stochastic quantum trajectories for disordered systems; (ii) a variational neural network method. The stochastic methods have been applied to investigate the driven-dissipative Bose-Hubbard lattice model that can for instance describe lattices of coupled nonlinear electromagnetic resonators such those obtained with semiconductor microcavities or superconducting quantum circuits. Our study has shown that a critical slowing down of the dynamics occurs for some parameters in dimensions higher than one. The phenomena is linked to a first-order dissipative phase transition and to lattice-level collective bistability. Furthermore, we reveal the role of disorder on the spatial correlation properties of these phenomena. In the last part of the thesis, we present the strong connection between fundamental concepts of machine learning and variational approaches in quantum physics. We present a novel computational method for open quantum systems based on neural network ansätze that is compatible with a stochastic minimisation of a cost function through Monte Carlo sampling. As a first application, we have applied this method to the dissipative transverse quantum Ising model. Exciting perspectives and possible developments are also drawn in the concluding remarks.

Keywords: Open Quantum Systems, Phase transitions, Quantum optics, Many-body physics, Neural Networks, Numerical methods, Monte Carlo, Disorder

Résumé: Cette thèse est consacrée à l'étude théorique des systèmes quantiques dissipatifs à plusieurs corps avec un accent particulier sur les méthodes de calcul avancées. Après une revue des plusieurs méthodes numériques de pointe pour déterminer l'état d'équilibre et la dynamique des systèmes quantiques ouverts décrits par une équation maitresse de Lindblad pour la matrice de densité, ce manuscrit présente deux méthodes originales: (i) un sampling optimal des trajectoires stochastiques quantiques pour les systèmes désordonnés; (ii) une méthode variationnel base sur les réseau neuronal. Les méthodes stochastiques ont été appliquées pour étudier le modèle de réseau de Bose-Hubbard dissipatif entraîné qui peut par exemple décrire des réseaux de résonateurs électromagnétiques non linéaires couplés tels que ceux obtenus avec des microcavités à semi-conducteurs ou des circuits quantiques supraconducteurs. Notre étude a montré qu'un ralentissement critique de la dynamique se produit pour certains paramètres de dimensions $D > 1$. Le phénomène est lié à une transition de phase dissipative du premier ordre et à une bistabilité collective au niveau du réseau. De plus, nous révélons le rôle du désordre sur les propriétés de corrélation spatiale de ces phénomènes. Dans la dernière partie de la thèse, nous présentons le lien étroit qui existe entre les concepts fondamentaux de l'apprentissage automatique et les approches variationnelles en physique quantique. Nous présentons une nouvelle méthode de calcul pour les systèmes quantiques ouverts basée sur une méthode de réseau de neurones compatible avec une minimisation stochastique d'une fonction de coût par échantillonnage de Monte Carlo. Comme première application, nous avons appliqué cette méthode au modèle Ising quantique transverse dissipatif. Des perspectives intéressantes et des évolutions possibles sont également exposées dans les remarques finales.

Mots-cle: Systèmes quantiques ouverts, Transitions de phase, Désordre, Optique quantique, Physique many-body, Réseaux de neurones, Méthodes numériques, Monte Carlo

List of Publications

During my Ph.D. I had the pleasure of publishing several works, which are listed below. This thesis focuses more on the first three.

[1] **F. Vicentini**, F. Minganti, R. Rota, G. Orso, and C. Ciuti,
Critical slowing down in driven-dissipative Bose-Hubbard lattices
Physical Review A **97**, 013853 (2018)

[2] **F. Vicentini**, F. Minganti, A. Biella, G. Orso, and C. Ciuti,
Optimal stochastic unraveling of disordered open quantum systems: Application to driven-dissipative photonic lattices
Physical Review A **99**, 032115 (2019)

[3] **F. Vicentini**, A. Biella, N. Regnault, and C. Ciuti,
Variational Neural-Network Ansatz for Steady States in Open Quantum Systems
Physical Review Letters **122**, 250503 (2019)

[4] V. Goblot, B. Rauer, **F. Vicentini**, A. Le Boité, E. Galopin, Lemaître, L. Le Gratiet, A. Harouri, I. Sagnes, S. Ravets, C. Ciuti, A. Amo, J. Bloch,
Discrete nonlinear domains for polariton fluids in a flat band
Physical Review Letters **123**, 113901 (2019)

[5] G. Carleo, K. Choo, D. Hofmann, J. E. T. Smith, T. Westerhout, F. Alet, E. J. Davis, S. Efthymiou, and I. Glasser, S. H. Lin, M. Mauri, G. Mazzola, C.B. Mendl, E.v. Nieuwenburg, O. O'Reilly, H. Théveniaut, G. Torlai, **F. Vicentini**, A. Wietek,
NetKet: A Machine Learning Toolkit for Many-Body Quantum Systems
SoftwareX **10**, 100311 (2019)

Contents

CHAPTER 1

Introduction

In recent years a significant amount of resources has been dedicated to the development of quantum technologies, which includes, among several, quantum computing and simulation [6]. With the development of science and technology, the need of developing tools to attack and solve some computationally-demanding problems has emerged. In general, it is possible to separate those computational problems into two classes: P-class problems which can be solved in deterministic polynomial time by ordinary calculators, while $NP-$class cannot [7, 8]. Advances in computer science produced several algorithms to solve $P-$problems in increasingly efficient ways. However it has long been understood that the asymptotical cost of NP problems cannot be made polynomial and therefore manageable [8]. Unfortunately, several problems that might guarantee interesting developments in medicine, material science and artificial intelligence belong to the NP class [9, 10, 11, 12, 13].

At the beginning of the 1980s physicists began to realise that it might be possible to exploit the fundamental properties of quantum mechanics to solve *some* of those problems in polynomial time. In particular, one might engineer a device that does not perform boolean operations on bits, but unitary operations on some fundamental quantum-informational unit, the qubit, realising a Digital Quantum Computer [14]. Alternatively, it might be possible to encode certain problems into the ground-state properties of a quantum-mechanical system, measurable in the laboratory, building an analog computer, known also as a quantum simulator [15, 16].

To build those devices one must ensure that the quantum nature of the devices is preserved during the computation. Fighting the inevitable relaxation to thermal equilibrium that destroys quantum information is traditionally achieved by properly isolating the device from the environment. Yet, despite the tremendous experimental progress in the isolation of quantum systems, a finite coupling to the environment is unavoidable and certainly plays a crucial role in the practical implementation of quantum information and quantum simulation protocols [17]. Building a computer capable of performing quantum calculations of arbitrary complexity would require an improvement of several orders of

magnitude in the quality of our devices. A paradigmatic shift in the error-correcting protocols would be required to account for qubit-level errors that inevitably happen during the calculation [18]. A simple theoretical description of those systems can be obtained within the realm of the traditional equilibrium quantum mechanics for closed systems which assumes that the degrees of freedom are decoupled from the rest of the universe, so that the resulting time-evolution is unitary. Such a strong approximation simplifies the theoretical description, but as we push our models to ever-longer time scales the description of quantum systems as isolated shows its limits. Quantum systems typically decay towards thermal states described by generalised Gibbs ensembles [19, 20], and their relaxation process can be modelled through unital channels with relative ease, but those tools are not able to capture the more general behaviour of non-equilibrium states that can be created by engineering the driving and the dissipation of such systems.

A tool to better confront those setups is the framework of *open quantum systems*, which allows for a more refined model of the interaction between the system and the environment, capturing non-equilibrium phenomena not treated within the traditional thermodynamical description. Historically the field was born out of the necessity of describing out-of-equilibrium quantum systems such as those found in cavity quantum electrodynamics setups. In those systems, one or more atoms interact with the quantum-electromagnetic field confined between two highly-reflective mirrors. Excited photons, however, inevitably leak outside of the mirrors, and the confined field must be continuously repopulated by optical pumping. Remarkably, experimentalists gather information about the system only through those photons that leak outside the system, and an understanding of the dissipation allows to model the observations.

The competition between the dissipation and the pump creates interesting novel phenomena, where thermodynamical equilibrium can be replaced by an effective dynamical equilibrium once the system has relaxed to its steady-state [21, 22]. These phenomena are common to several experimental platforms exploiting light-matter interaction, such as superconduting quantum circuits [23, 24], semiconductor cavities [25, 26] and optomechanical resonators [27]. Those systems are particularly interesting in the regime where the matter and wave degrees of freedom strongly couple, and are canonically described by means of hybrid bosonic quasi-particles called polaritons [26, 28]. Polaritons can be interpreted as photon-like particles is dressed by the matter degree of freedom, which can be represented by phonons [29], excitons [30, 31], plasmons [32] or intersubband excitations [33] to give a few examples. Even though photons in vacuum are known to interact so weakly that a photon-photon interaction event has only been recently detected in high-energy experiments [34], polaritons can be engineered to have very strong interactions, while retaining the controllability of optical frequencies. Those properties make polaritons a good platform to simulate bosonic Hamiltonians or to use as the foundation of a Quantum Computer architecture [35], but one must first understand the fundamental laws that govern such dissipative systems.

While at equilibrium interesting physical properties such as Bose-Einstein condensation emerge when many particles are made to interact [36], and they can be understood through general thermodynamical paradigms such as free energy and symmetry considerations, those tools do not generally work out of equilibrium and our understanding of what drives transitions is far from complete. For example, the fate of the famous Noether's theorem

on the link between symmetries and conserved quantities has only recently been extended to dissipative systems [37, 38], as well as the physical interpretation of the spectrum of those systems [39, 40] and the emergence of criticality in presence of phase transitions does not follow the same paradigms of equilibrium statistical mechanics [41, 42, 43, 44, 45, 46]. Recently, it has also been shown that engineering non-thermal, parametric baths creates non-trivial attractors [47, 48, 49, 50] which can in some cases enforce dissipative symmetries in the non-equilibrium steady-state, a promising candidate for dissipation-resilient qubit encodings [51].

Understanding the fundamental laws that govern non-equilibrium quantum systems is therefore important to the development of science and quantum technologies. However, the theoretical study of such systems is hampered by the remarkable complications that arise in this context. In particular, there exist only few systems for which one can compute analytically observables in the steady-state [52, 53, 54], and non-observables like operator-entanglement or entropy cannot be computed, while the analytical solution to the full dynamics can be recovered only for a handful of almost-linear transport models that can be fully solved if dissipation is added at the edges of a chain [55] thanks to a general technique known as third-quantisation [56].

For generic systems that do not respect those conditions, researchers usually resort to a combination of several approximation schemes as well as numerical simulations. Yet, the exponential growth of the Hilbert space with the system size combined with the need of representing mixed states [21] makes the simulation of open quantum many-body systems a formidable task. It is for that reason that, researchers have devoted significant resources both to the study of fundamental properties of dissipative systems and to the development of powerful computational methods to simulate the physics of open systems. Fruitful developments came from the generalisation of established techniques originally developed for closed systems, such as renormalisation group calculations in the Keldysh formalism [57, 58], impurity solvers for dynamical mean field theory [59], permutation-invariant solvers [60], the full configuration-interaction Monte Carlo [61] and cluster methods [62, 63]. Several complications have been encountered when trying to generalise Tensor-Network techniques to dissipative systems [64, 65, 66], and different solutions to the problem of positive-definite evolution have been proposed [67, 68]. Lastly, there also exist other numerical techniques that have been developed natively for dissipative systems, such as the corner-space renormalisation method [69].

Having noticed the effectiveness of neural-networks in solving generic optimisation problems, several groups have tried to bridge the gap between the language of machine learning and numerical physics. This effort resulted in a series of remarkable results both for equilibrium [70, 71, 72] and dissipative systems [3, 73, 74, 75]. Nevertheless bringing closer together machine learning and quantum physics has not only been useful to produce novel numerical methods, but it has also produced interesting algorithms integrating to perform machine-learning of quantum circuits [76, 77], thanks to quantum gate automatic differentiation [78, 79] and to classical or quantum optimisation procedures of potentially NP problems [80, 81].

This thesis presents the main results obtained during the three years of my Ph.D.

research in Physics. In Chapter 2 we introduce the theoretical tools needed to describe open quantum systems. In Section 2.1 we introduce the Lindblad master equation and give an interpretation in terms of continuous measurement theory, while in Section 2.2 we give an overview of the unraveling of quantum trajectories.

In Chapter 3 we summarise several numerical methods developed to treat those systems, beginning with generic linear solvers in Section 3.1, steady-state methods (3.2.1), and tensor networks (3.3). In Section 3.4 we comprehensively discuss trajectory methods that will be used in the following chapter, and in Section 3.5 we introduce our own sampling scheme to deal with disordered quantum systems (based on [2]).

Chapter 4 is devoted to the understanding of a generic first-order dissipative phase transition. In particular, the dynamical properties are studied numerically in Section 4.3 (based on [1]) and the statical properties as well as the effect of disorder are discussed in Section 4.4 (based on [3]).

In Chapter 5 we provide an introduction to neural-networks in order to later introduce our own novel numerical method to treat open quantum systems, based on [3]. Future developments are discussed at the end of the chapter.

Finally, we conclude the manuscript with Chapter 6, where we summarise our conclusions discussing the possible future directions.

CHAPTER 2

Theory of open quantum systems

Any quantum system cannot be perfectly isolated from the environment, even more so if we wish to measure it. It is therefore important to work within a framework where the processes through which the system interacts with the environment are taken into account. Remarkably, the action of the environment onto the system can be understood in terms of a continuous measurement. In this chapter we will first introduce the theoretical framework of the Lindblad Master Equation (Section 2.1), which describes the time evolution of quantum systems coupled to memoryless environments, and provide the two complementary interpretations in terms of continuous monitoring (Section 2.1.2) and starting from microscopic principles (Section 2.1.3). We then introduce the Stochastic Schrödinger Equation (Section 2.2), which provides an interpretation of the Lindblad Master Equation in terms of stochastic processes. In Section 2.3 we revisit a technique to reformulate the master equation in phase space, which will later be used to derive efficient numerical approximations.

Contents

2.1 The Lindblad master equation

As we have discussed in the introduction, a quantum system cannot be perfectly isolated from the experimental apparatus and from its environment. Modelling the effect that this interaction has on the dynamics is not an easy task, as the environment is in principle an infinitely large system, in itself already intractable and difficult to describe [21, 22, 82]. Keeping track of the precise microscopic state of the environment, and therefore its action at all times, is virtually impossible. Therefore we cannot hope to have a perfect knowledge of the state of the system, and we must resort to a statistical description of the quantum system. In this section we briefly introduce the required theoretical tools.

2.1.1 Density matrix and quantum maps

In quantum mechanics the state of an isolated quantum system S with Hilbert space $\mathcal{H}_S$ is described by a state vector or wavefunction $|\psi\rangle \in \mathcal{H}_S$. Quantities of interest describing the state of the system are called *observables* and are described by self-adjoint operators $\hat{A} \in \mathcal{B}(\mathcal{H}_S)$ acting on the Hilbert space[1] [83]. In the canonical formulation of quantum mechanics, the wavefunction has a statistical meaning [84], and observables can be computed by means of the expectation values:

$$\langle \hat{A} \rangle = \langle \psi | \hat{A} | \psi \rangle \,. \tag{2.1}$$

The wavefunction is also called a pure-state, and encodes information about uncertainities arising from quantum-mechanical properties of the physical system.

Density Matrices

When dealing with finite temperature or systems coupled to an environment a quantum system is, in general, not in a pure state. For example a spin system might be in the state $|\psi_\uparrow\rangle$ with probability $p_\uparrow \approx 1$ but the orthogonal state $|\psi_\downarrow\rangle$ can occur with probability $p_\downarrow = 1 - p_\uparrow$. In such cases, we can account for this additional uncertainty by performing a weighted sum over all possible pure states, namely,

$$\langle \hat{A} \rangle = \sum_i p_i \langle \psi_i | \hat{A} | \psi_i \rangle \,. \tag{2.2}$$

We stress the difference between the quantum-related uncertainty, arising from the intrinsic quantum expectation value over pure states, and the classical uncertainty arising from the average over the classical ensemble $\{\psi_\downarrow, \psi_\uparrow\}$.

[1] $\mathcal{B}(\mathcal{H})$ is the space of bounded linear operators acting on $\mathcal{H}$. If $A \in \mathcal{B}(\mathcal{H}) \to B : \mathcal{H} \to \mathcal{H}$, then there exists an $M \geq 0$ such that, $\forall\, |\psi\rangle \in \mathcal{H}$, $\langle \psi | A | \psi \rangle \leq M \langle \psi | \psi \rangle$.

Mathematically, the object encoding all the statistical properties of a classical mixture is the covariance matrix of a random variable [85, 21]. In the case where the random variable is a pure quantum state $|\psi\rangle$, the covariance matrix has the form

$$\hat{\rho} = \mathbb{E}[|\psi\rangle \langle\psi|] = \sum_i p_i |\psi_i\rangle \langle\psi_i| , \tag{2.3}$$

and is called the density matrix. The density matrix encodes a *mixed state*, which is an incoherent mixture of pure quantum states[2]. We recall that the density matrix object has the following properties:

- Trace 1, $\sum_i p_i = 1$ by construction, to normalize probability;
- $\hat{\rho}$ is hermitian ($\hat{\rho} = \hat{\rho}^\dagger$);
- $\hat{\rho}$ is positive-semidefinite, that is, $\forall\psi \Rightarrow \langle\psi|\hat{\rho}|\psi\rangle \geq 0$;
- If the state is mixed, there exist infinitely many decompositions like Eq. (2.3), not all of which are orthogonal. Otherwise, if the state is pure, the decomposition is unique and there is only one wavefunction.

Quantum Maps

Having defined the mathematical object encoding mixed states in the previous section, a general problem of interest is that of describing its time evolution. In general, the time-evolution of the density matrix operator $\hat{\rho}(t)$ will be generated by the so-called quantum map $\Phi : \mathcal{B}(\mathcal{H}_S) \to \mathcal{B}(\mathcal{H}_S)$ according to

$$\hat{\rho}(t + dt) = \Phi(\hat{\rho}(t)). \tag{2.4}$$

The quantum map Φ is a *superoperator*, because it acts on operator spaces. As the quantum map evolves quantum states, it must preserve the properties of the density matrix described above, therefore it:

- must be linear, $\Phi(\alpha\hat{\rho}_a + \beta\hat{\rho}_b) = \alpha\Phi(\hat{\rho}_a) + \beta\Phi(\hat{\rho}_b)$;
- must preserve the hermitian nature of $\hat{\rho}$, therefore if $\hat{\rho} = \hat{\rho}^\dagger \Rightarrow \Phi(\hat{\rho}) = (\Phi(\hat{\rho}))^\dagger$;
- must preserve the positive definit character of $\hat{\rho}$, namely $\forall\psi \Rightarrow \langle\psi|\Phi(\hat{\rho})|\psi\rangle \geq 0$.
- must preserve the trace, i.e. $Tr[\hat{\rho}] = \mathrm{Tr}[\Phi(\hat{\rho})]$;

Maps that respect those 4 properties are known as *completely positive trace preserving maps* (CPTP). Kraus theorem states that CPTP maps can be always decomposed in the form

$$\Phi(\hat{\rho}) = \sum_k \hat{B}_k \hat{\rho} \hat{B}_k^\dagger, \tag{2.5}$$

[2]By incoherent mixture we mean, in statistical terms, a classical superposition. In the following the two definitions will be used interchangeably.

where the set of operators $\hat{B}_k \in \mathcal{B}(\mathcal{H}_S)$ satisfy the condition[3] $\sum_k \hat{B}_k^\dagger \hat{B}_k = \mathbb{I}$ [86]. This decomposition is also known as the *Kraus decomposition* of a quantum map.

In general, it is useful not to study directly the quantum map, but the differential superoperator that generates the map according to the equation

$$\frac{d\hat{\rho}}{dt} = \mathcal{L}(\hat{\rho}). \tag{2.6}$$

$\mathcal{L}(\hat{\rho})$ is often called the Liouvillian because of an analogy between the equation above and the Liouville equation of classical mechanics. The quantum map is then obtained by formally integrating over time the above equation.

Physical quantum maps are completely-positive and trace-preserving, therefore also the Liouvillian generating it must have the same properties. Remarkably, Lindblad proved in 1976 that the Liouvillian can always be decomposed into the following form[4] (setting $\hbar = 1$, which will be assumed for the rest of the manuscript):

$$\mathcal{L}(\hat{\rho}) = -i\Big[\hat{H}, \hat{\rho}\Big] + \sum_i \left(\hat{L}_i \hat{\rho} \hat{L}_i^\dagger - \frac{1}{2}\Big\{\hat{\rho}, \hat{L}_i^\dagger \hat{L}_i\Big\}\right), \tag{2.7}$$

where H is an Hermitian operator and L_k are arbitrary *jump* operators [87, 88]. This form is commonly called the Lindblad master equation, or, more generally, Gorini-Kossakowski-Sudarshan-Lindblad (GKSL) equation. Nevertheless, this is a purely mathematical result, and care must be taken to obtain a physical interpretation of the various terms above, which will be done in the following paragraphs.

2.1.2 Continuous monitoring interpretation

In the previous paragraph we have seen that the time-independent physical evolution of a quantum system in a mixed state must be of the form given by Eq. (2.27). In this section we would like to give an interpretation of the above equation in terms of a system being continuously measured. To do so, we will first briefly introduce the formalism of generalised Positive-Operator Valued Measures (POVM), with which generalised measurements (non self-adjoint) on mixed states are represented.

Measurements on mixed states

In general, a measurement with possible outcomes $\{r\}$ can be completely specified in terms of the set of measurement operators $\{\hat{M}_r\}$. The probability to measure the outcome r is given by the expectation value of the operator $\hat{E}_r = \hat{M}_r^\dagger \hat{M}_r$. Therefore to normalise the probability, the operators must satisfy the completeness condition

$$\sum_r \hat{M}_r^\dagger \hat{M}_r = \mathbb{I}. \tag{2.8}$$

[3] For a completely positive map, which does not preserve (therefore contracts) the trace, the normalization condition can be relaxed to $\sum_k \hat{B}_k^\dagger \hat{B}_k = \alpha \mathbb{I}$ for $\alpha \leq 1$.

[4] More technically, Lindblad's theorem states that *The generator of any quantum operation satisfying the semigroup property must have the form of Eq.* (2.27). In any case, one can show that any quantum operation generated by exponentiating a time-independent superoperator satisfy the semigroup property.

Upon performing the measurement $\{\hat{M}_r\}$ on $\hat{\rho}$, the outcome r will be obtained with probability $p_r = \mathrm{Tr}\left[\hat{M}_r\hat{\rho}\hat{M}_r^\dagger\right]$, after which the state will be

$$\hat{\rho}_r = \frac{\hat{M}_r\hat{\rho}\hat{M}_r^\dagger}{\mathrm{Tr}\left[\hat{M}_r\hat{\rho}\hat{M}_r^\dagger\right]}, \tag{2.9}$$

which is the properly normalised covariance matrix of $\hat{M}_r|\psi\rangle$[5]. If a measurement is carried out on a state $\hat{\rho}$, but the result is unknown, or not kept track of, we say that the measurement was *unread*. The state $\hat{\rho}'$ after an unread measurement is the sum of all possible outcomes $\hat{\rho}_r$ weighted by their respective probabilities p_r,

$$\hat{\rho}' = \sum_r p_r\hat{\rho}_r = \hat{M}_r\hat{\rho}\hat{M}_r^\dagger. \tag{2.10}$$

This form is the same of Eq. (2.5), therefore the act of measuring a system without reading the result generates a valid CPTP quantum map.

Quantum maps as continuous measurements

To interpret the Lindblad master equation 2.27 in terms of measurement operators, we start by writing the coherent evolution of an isolated quantum system with Hamiltonian $\hat{H}$ in terms of the Kraus decomposition defined by Eq. (2.5). The following ansatz,

$$\hat{B} = \mathbb{I} - idt\hat{H}, \tag{2.11}$$

respects the normalization condition $\hat{B}^\dagger\hat{B} = \mathbb{I}$ up to first order in dt and generates the quantum map,

$$\begin{aligned}
\hat{\rho}(t+dt) &= \Phi(\hat{\rho}(t)) && (2.12)\\
&= (\mathbb{I} - idt\hat{H})\hat{\rho}(t)(\mathbb{I} + idt\hat{H}) && (2.13)\\
&= \hat{\rho}(t) - idt[\hat{H}, \hat{\rho}(t)] + \mathcal{O}(dt^2). && (2.14)
\end{aligned}$$

If the time interval is small enough ($dt \to 0$), this map is equivalent to the Heisenberg equation of motion,

$$\frac{d\hat{\rho}(t)}{dt} = \frac{\hat{\rho}(t+dt) - \hat{\rho}(t)}{dt} = -i\left[\hat{H}, \hat{\rho}(t)\right]. \tag{2.15}$$

We now seek to generalise this unitary evolution by incorporating measurements according to an operator $\hat{L}$ with two possible outcomes, $r = \{0, 1\}$. To preserve the normalisation condition $\sum_r \hat{B}_r^\dagger\hat{B}_r = \mathbb{I}$, a set of at least two Kraus operators will be necessary to describe the dynamics, namely:

$$\begin{cases} \hat{B}_0 = \mathbb{I} - idt(\hat{H} - i\hat{L}^\dagger\hat{L}/2) \\ \hat{B}_1 = \sqrt{dt}\hat{L} \end{cases} \tag{2.16}$$

[5] Assuming ψ is a random variable as defined in Section 2.1

If the system evolving under Hamiltonian $\hat{H}$ is continuously measured with $\hat{L}$ but the result of the measurement is unread, the equation of motion is obtained by substituting Eqs. (2.10) and (2.39) into Eq. (2.15), which yields

$$\frac{d\hat{\rho}(t)}{dt} = \frac{\hat{B}_0\hat{\rho}(t)\hat{B}_0^\dagger + \hat{B}_1\hat{\rho}(t)\hat{B}_1^\dagger - \hat{\rho}(t)}{dt}, \tag{2.17}$$

$$= -i(\hat{\tilde{H}}\hat{\rho}(t) - \hat{\rho}(t)\hat{\tilde{H}}^\dagger) + \hat{L}\hat{\rho}(t)\hat{L}^\dagger, \tag{2.18}$$

where $\hat{\tilde{H}} = \hat{H} - i\hat{L}^\dagger\hat{L}/2$ is the effective non-hermitian Hamiltonian. This equation is equivalent to the Lindblad master equation (2.27), but now the $\hat{L}$ operators have a precise physical meaning: they represent a continuous binary measurement being performed on the system.

The probability to have the outcome $r = 1$ in the measurement during an infinitesimal time interval dt is given by the infinitesimal expectation value

$$\mathcal{P}_1 = dt\,\mathrm{Tr}\Big[\hat{L}^\dagger\hat{L}\hat{\rho}\Big]. \tag{2.19}$$

The probability for the measurement $r = 0$, $\mathcal{P}_0 = 1 - \mathcal{O}(dt)$, is of the order of unity and therefore it is regarded as a null result. If the measurement gives the null result, the system state changes infinitesimally, but not unitarily, according to the evolution generated by the non hermitian Hamiltonian $\hat{\tilde{H}}$. At random times distributed according to the rate $\mathcal{P}_1/dt$, the measurement gives the result $r = 1$, and the system undergoes a finite evolution induced by the operator $\hat{B}_1$, which we call a *quantum jump*.

Comparing the first term of Eq. (2.39) with the coherent evolution generated by Eq. (2.11), it is clear that the non hermitian operator $\hat{\tilde{H}} = \hat{H} - i\hat{L}^\dagger\hat{L}/2$ plays the same role of the continuous time evolution of the Hamiltonian in the coherent case. In the context of non-hermitian (or more generally PT-preserving) quantum mechanics only the dynamics generated by the non hermitian Hamiltonian $\hat{\tilde{H}}$ is considered, and the term generating quantum jumps is neglected.

Lastly, we remark that this proof can be generalised to the case of an arbitrary (finite) number N of measurement operators $\hat{L}_i$. In this case the decomposition involves the $N+1$ Kraus operators

$$\begin{cases} \hat{B}_0 = \mathbb{I} - idt(\hat{H} - i\sum_i \hat{L}_i^\dagger\hat{L}_i/2) \\ \hat{B}_i = \sqrt{dt}\hat{L}_i \end{cases}, \tag{2.20}$$

which can be interpreted as before, and the differential equation governing the dynamics is in the Lindblad form, where the set of operators $\{\hat{L}_i\}$ are the unread measurements performed on the system.

2.1.3 Microscopic interpretation

We have shown in the previous section that the Lindblad master equation encodes the evolution of a quantum system subject to continuous measurement by an observer. As we will see, the Master Equation formalism is more general, as it also describes the stochastic evolution of *open quantum systems*, that is, systems not perfectly isolated from the

environment. In the following, we sketch a microscopic derivation of the Lindblad master equation in terms of a system coupled to a reservoir largely following the derivation of Breuer and Petruccione [21]. This derivation will be useful to interpret the coupled environment as an observer measuring the quantum system.

The reduced density matrix

In general, the state of the quantum system S is a wave-function in the Hilbert space $\mathcal{H}_S$, and the environment E has a state defined in an infinitely large space $\mathcal{H}_E$, obtained by coupling an infinite number of degrees of freedom together. The wavefunction $\Psi \in \mathcal{H}_{SE}$ describing the system and environment will be a vector in the *"universe"* Hilbert space $\mathcal{H}_{SE} = \mathcal{H}_S \otimes \mathcal{H}_E$. This object holds precise informations not only about the system, but also about correlations in the environment. To reduce the complexity of the problem, we can forgo the exact description in terms of pure states of the quantum system [21, 89], and describe its state as a statistical mixture Section 2.1.1. The statistical uncertainty arises from our ignorance of the precise state of the environment.

To obtain the reduced density matrix for the system S starting from a pure state Ψ, one can perform the partial trace of the universe density matrix $|\Psi\rangle\langle\Psi|$,

$$\hat{\rho}_S = \mathrm{Tr}_E[|\Psi\rangle\langle\Psi|] = \sum_{ij} c_{i,j} |\psi_i\rangle\langle\psi_j| \tag{2.21}$$

where $|\psi_i\rangle \in \mathcal{H}_S$ is a pure state of the system. The reduced density matrix is, to all effects, a density matrix and has the same properties that were introduced in Section 2.1.1.

The master equation

We now wish to derive an expression of the quantum map evolving the reduced density matrix $\hat{\rho}_S$. We recall that Lindblad's theorem states that this quantum map must have the Lindblad form Eq. (2.27), but the derivation will allow us to give a precise microscopic interpretation of the measurement operators $\hat{L}$.

We assume that the pure-state $|\Psi\rangle$ of the universe evolves coherently under the Hamiltonian dynamics $\hat{H}$

$$\hat{H} = \hat{H}_S \otimes I_E + I_S \otimes \hat{H}_E + \hat{H}_I, \tag{2.22}$$

where $\hat{H}_S \in \mathcal{B}(\mathcal{H}_S)$ is the Hamiltonian of the system, $\hat{H}_E \in \mathcal{B}(\mathcal{H}_E)$ is the Hamiltonian of the environment and $\hat{H}_I \in \mathcal{B}(\mathcal{H}_S \otimes \mathcal{H}_E)$ is the interaction Hamiltonian. $I_{S/E}$ is the identity operator acting on the system and environment Hilbert spaces respectively. To derive the equations of motion for the system's density matrix, it is useful to assume that the coupling between the system and the reservoir is weak (this is also known as the *Born approximation*).

Thus, the density matrix of the reservoir is only negligibly affected by the interaction and the state at all times t may be approximately by a tensor product

$$\hat{\rho}(t) \approx \hat{\rho}_S(t) \otimes \hat{\rho}_B, \tag{2.23}$$

where we also assumed that the reservoir is in an equilibrium state $\hat{\rho}_B$ that is no longer evolving due to its internal dynamics. We will assume the rather general form of interaction

$$\hat{H}_I = \sum_i \hat{L}_i \otimes \hat{B}_i \tag{2.24}$$

where $\hat{L}_i \in \mathcal{H}_S$ is acting on the system only and $\hat{B}_i$ on the bath only.

A further approximation is assumption of Markovian dynamics. This means that the instantaneous action depends only on the state at time t and not on its history. This is consistent with our assumption that the reservoir is made of a continuous collection of infinite degrees of freedom, and its time-correlation functions decay faster then the time scales of the system. Lastly, to obtain a valid equation we also need to perform the secular (or rotating wave) approximation, by integrating over rapidly decaying degrees of freedom.

Starting from the Heisenberg's equation for the universe density matrix,

$$\frac{d\left|\Psi\right\rangle\left\langle\Psi\right|}{dt} = -i\Big[\hat{H}, \left|\Psi\right\rangle\left\langle\Psi\right|\Big], \tag{2.25}$$

if we take the partial trace over the environment on both sides of the equation, we obtain the expression

$$\frac{d\hat{\rho}_S}{dt} = -i\Big[\hat{H}_S, \hat{\rho}_S\Big] + \mathrm{Tr}_E[-i[H_I, \hat{\rho}_S \otimes \hat{\rho}_B]]. \tag{2.26}$$

Using the Markov and secular approximations, it can be simplified into the Lindblad master equation

$$\frac{d\hat{\rho}_S}{dt} = -i\Big[\hat{H}_S, \hat{\rho}_S\Big] + \sum_i \left(\hat{L}_i\hat{\rho}_S\hat{L}_i^\dagger - \frac{1}{2}\Big\{\hat{\rho}_S, \hat{L}_i^\dagger\hat{L}_i\Big\}\right), \tag{2.27}$$

where $\hat{L}_i$ are the so-called jump operators describing internal transitions due to the interaction with the environment. We remark that if the system and the bath do not interact, $\hat{L} = 0$ and therefore this equation reduces to the standard Heisenberg equation of motion for the density matrix. It is possible to show that in some particular cases a non-markovian environment introduces a dependency of the system-bath coupling $\hat{H}_I$ on spectrum of the system. In this case the resulting master equation has the same form, and the frequency-dependence can be hidden inside of the jump operators $\hat{L}_i$ [90, 91].

Environment as an observer

We have shown that the dynamics of a quantum system weakly coupled to an environment, which acts on the system according to the operator $\hat{L}$, is equivalent to the dynamics describing the evolution of a quantum system under continuous unread measurements described by the same jump operator.

To give a more through interpretation, let us consider an example: a bosonic system with trivial dynamics $\hat{H} = 0$, linearly coupled with rate γ to a zero-temperature bath modelled with the jump operator $\hat{L} = \hat{a}$. In this case, the master equation is

$$\mathcal{L}\hat{\rho}(t) = \gamma\left[\hat{a}\hat{\rho}(t)\hat{a}^\dagger - \frac{1}{2}\Big\{\hat{\rho}(t), \hat{a}^\dagger\hat{a}\Big\}\right]. \tag{2.28}$$

As we have shown previously, this master equation is generated by the two Kraus Operators

$$\begin{cases} \hat{B}_0 = \mathbb{I} - \frac{dt\gamma}{2}\hat{a}^\dagger\hat{a} \\ \hat{B}_1 = \sqrt{dt\gamma}\ \hat{a} \end{cases} \tag{2.29}$$

and since they respect the trace condition $\hat{B}_0^\dagger\hat{B}_0 + \hat{B}_1^\dagger\hat{B}_1 = \mathbb{I}$, they form a generalised measurement (POVM) on the system.

The two outcomes of the measurements $r = \{0, 1\}$ are observed with the probabilities

$$\begin{cases} p_0(t) = \mathrm{Tr}\left[\hat{B}_0\hat{\rho}(t)\hat{B}_0\right] = 1 - dt\gamma\,\langle n\rangle\,(t) \\ p_1(t) = \mathrm{Tr}\left[\hat{B}_1\hat{\rho}(t)\hat{B}_1\right] = dt\gamma\,\langle n\rangle\,(t), \end{cases} \tag{2.30}$$

where $\langle n\rangle\,(t) = \mathrm{Tr}\left[\hat{a}^\dagger\hat{a}\hat{\rho}(t)\right]$. With probability of order 1 the outcome of the measurement is $r = 0$ (no signal), and the system is unchanged. With probability of order $dt\gamma$ and proportional to the number of photons inside the cavity the observed measurement is $r = 1$ (a signal) and the system is changed by applying the bosonic destruction operator to the system according to

$$\hat{\rho}(t + dt) = \frac{\hat{a}\hat{\rho}(t)\hat{a}^\dagger}{\mathrm{Tr}[\hat{a}\hat{\rho}(t)\hat{a}^\dagger]}. \tag{2.31}$$

This means that if the measurement gives a positive result, the measurement destroyed a photon in the system. Moreover, the rate at which the measurement destroys photons in the system is $\tau = p_1(t)/dt = \gamma\,\langle n\rangle\,(t)$. This allows us to interpret the action of the environment as an observer looking at photons coming out the system.

2.1.4 The Liouvillian spectrum and the steady-state

The Liouvillian can always be decomposed into its eigenvalues. Since it is not hermitian, in general left ($\hat{\rho}_\lambda$) and right ($\hat{\pi}_\lambda$) eigenstates do not coincide, and the eigenvalues λ are complex numbers. Left and right eigenstates satisfy the equations:

$$\mathcal{L}\hat{\rho}_\lambda = \lambda\hat{\rho}_\lambda \tag{2.32}$$

$$\hat{\pi}_\lambda\mathcal{L} = \lambda\hat{\pi}_\lambda \tag{2.33}$$

It can be proven that if a single eigenvalue is real, then left and right eigenstates coincide [40]. In the following, we want to show that the spectrum of a Liouvillian is closely related to the dynamics of the system. In particoular, by integrating over time the Lindblad master equation 2.27, it is possible to obtain the formal expression of the quantum map Φ as

$$\hat{\rho}(t) = e^{\mathcal{L}t}\hat{\rho}(0), \tag{2.34}$$

where $\hat{\rho}(0)$ is the initial state at $t = 0$. Assuming the Liouvillian is time-independent and the Hilbert space is finite-dimensional then there always exists at least one steady-state $\hat{\rho}_{ss}$ [21], which is defined as

$$\mathcal{L}\hat{\rho}_{ss} = 0. \tag{2.35}$$

Steady states of a Liouvillian are elements of its kernel, and therefore are 0-eigenvalues of the super-operator $\mathcal{L}$. The 0-eigenvalue has algebraic molteplicity equal to the number of

steady states. The emergence of multiple steady states has been associated by Albert et al. to super-symmetries [37]. As the invariance under a super-symmetry of the Liouvillian is a very strong requirement, and easy to verify analitically, in this thesis we will always assume that the systems studied have a single steady state. Regardless, the numerical methods developed in this thesis should be easily extensible to the case of multiple steady states by accurate decomposition of the operators involved.

Physically, the steady state is obtained by taking the infinite-time limit solution

$$\hat{\rho}_{ss} = \lim_{t\to\infty} \hat{\rho}(t) = \lim_{t\to\infty} e^{\mathcal{L}t}\hat{\rho}(0). \tag{2.36}$$

If the system has a single steady state, it will eventually converge to it, regardless of the initial conditions [92]. If the system has multiple steady-states, taking the infinite-time limit gives in general a linear combination of the steady states [37, 38].

The dynamics of the density matrix can be expanded into the eigenstates of the Liouvillian:

$$\hat{\rho}(t) = \hat{\rho}_{ss} + \sum_{i\neq 0} c_i(0) e^{\lambda_i t}\pi_i. \tag{2.37}$$

From this equation, it is clear that the real part (which is always negative) of the eigenvalues govern the relaxation towards the steady state, while the imaginary part gives an oscillatory behaviour, that will eventually disappear due to the damping. As the real part of the eigenvalues sets the damping rate, it is customary to order the eigenvalues in such a way that $\lambda_0 < |\mathbb{R}e[\lambda_1]| < |\mathbb{R}e[\lambda_n]$. This way the first few eigenvalues determine the long-time dynamics. Assuming that a single steady-state exists, the asymptotic behaviour in the long-time limit is given by

$$\hat{\rho}(t\to\infty) \simeq \hat{\rho}_{ss} + e^{-|\lambda_1|t}\pi_1. \tag{2.38}$$

$\mathbb{R}e[\lambda_1]$ is usually called the Liouvillian gap or the asymptotic decay rate [39].

2.2 Quantum trajectories

The Lindblad master equation introduced in the previous section encodes the dynamics of a quantum system subject to continuous measurements onto an equation of motion for its reduced density matrix $\hat{\rho}_S(t)$. This object represents a classical mixture of quantum states defined on the underlying Hilbert space $\mathcal{H}_S$, which can be interpreted as a functional $P[\psi] : \mathcal{H} \to \mathbb{R}$, associating a probability to states. To make an analogy with classical stochastic processes we recall that the Fokker-Planck equation is also an equation of motion for a probability functional, describing the dynamics of a whole ensemble of particles evolving according to a stochastic process. In similar spirit, in this section we want to show how it is possible to *unravel* the ensemble dynamics of the Liouvillian $\mathcal{L}$ into a stochastic process $\psi(t)$ defined over the system's Hilbert space $\mathcal{H}_S$. The stochastic process encodes the evolution of a single element of the classical ensemble, which we call a *quantum trajectory*. The quantum trajectories are obtained by recording the outcomes of the measurements carried out by the environment, instead of performing an average over all possible results as done in Section 2.1.2.

While the point of view of trajectories and stochastic processes are consistent with the ensemble dynamics after taking an ensemble and noise average, more information can be extracted by looking at their statistical properties, as will be shown in Section 4.3.

The mathematical remapping of the master equation into a stochastic process was originally proposed by Davies [93], and was only later given a physical interpretation [94]. The treatment presented here is largely inspired by Breuer and Petruccione [21], Gardiner and Zoller [85], Carmichael [95], the seminal paper by Dalibard, Castin and Molmer [96] and by a few other references [97].

2.2.1 The stochastic Schrödinger equation

We recall that a mixed state $\hat{\rho} = \sum_i p_i \left|\psi\right\rangle \left\langle\psi\right|$ represents a classical mixture of several pure quantum states. The uncertainty arises from the fact that the measurement performed by the environment is unread. If one records the result of every measurement, which is a stochastic process, it is possible determine the precise evolution of the quantum state. We consider a system initialized in a pure state $|\psi(t)\rangle$, evolving under the coherent dynamics of the Hamiltonian $\hat{H}$, and subject to the measurement (dissipation) by the operator $\hat{L}$. As we have shown before, the action of the dissipation is equivalent to a continuous unread measurement according to the two measurement operators

$$\begin{cases} \hat{B}_0 = \mathbb{I} - idt(\hat{H} - i\hat{L}^\dagger\hat{L}/2), \\ \hat{B}_1 = \sqrt{dt}\hat{L}. \end{cases} \tag{2.39}$$

The result of the measurement at each infinitesimal time increment dt is a random poissonian process $dr(t) = \{0, 1\}$. At every timestep, the expectation value of $dr(t)$ is the probability to detect $r = 1$, which is given by

$$\mathbb{E}[dr(t)] = dt \left\langle\psi(t)|\hat{L}^\dagger\hat{L}|\psi(t)\right\rangle . \tag{2.40}$$

At every time t, if $dr(t) = 0$ the measurement gives the null result and therefore the system undergoes a smooth non unitary evolution induced by the non hermitian Hamiltonian,

$$|\psi_0(t+dt)\rangle = \frac{\hat{B}_0\left|\psi(t)\right\rangle}{\sqrt{\left\langle\psi(t)|\hat{B}_0^\dagger\hat{B}_0|\psi(t)\right\rangle}} \tag{2.41}$$

$$= \left(1 - idt\left[\hat{\tilde{H}} + i\frac{1}{2}\left\langle\psi_0(t)|\hat{L}^\dagger\hat{L}|\psi_0(t)\right\rangle\right] + \mathcal{O}(dt^2)\right)\left|\psi(t)\right\rangle . \tag{2.42}$$

On the contrary, if the result of measurement is $dr(t) = 1$ the measurement is positive and the state changes significantly following a quantum jump. After the jump the state becomes, to first order in dt,

$$|\psi_1(t+dt)\rangle = \frac{\hat{B}_1\left|\psi(t)\right\rangle}{\sqrt{\left\langle\psi(t)|\hat{B}_1^\dagger\hat{B}_1|\psi(t)\right\rangle}} = \frac{\hat{L}\left|\psi(t)\right\rangle}{\sqrt{\left\langle\psi(t)|\hat{L}^\dagger\hat{L}|\psi(t)\right\rangle}}, \tag{2.43}$$

By combining the two evolutions, we can write the wavefunction $|\psi(t)\rangle$ as a stochastic process, the so-called Stochastic Schrödinger Equation (SSE),

$$d|\psi(t)\rangle = [(1 - dr(t))|\psi_0(t + dt)\rangle + dr(t)(|\psi_1(t + dt)\rangle - |\psi(t)\rangle)] \tag{2.44}$$

$$= \left[idt \left(-\hat{\tilde{H}} - i\frac{1}{2} \langle \hat{L}^\dagger \hat{L} \rangle_{\psi(t)} \right) + dr(t) \left(\frac{\hat{L}}{\sqrt{\langle \hat{L}^\dagger \hat{L} \rangle_{\psi(t)}}} - 1 \right) \right] |\psi(t)\rangle \tag{2.45}$$

where we used the property of stochastic calculus that $dr(t)dt = 0$ [85, 82, 98]. By closely inspecting Eq. (2.45) we see that its first term closely resembles that of the Schrödinger's equation, while the second term is a jump term. Equations of this form are jump-diffusion equations, which are a particular type of Stochastic Differential Equations (SDE).

To show that this process is consistent with the ensemble evolution of the Lindblad master equation, one simply needs to compute the covariance matrix of this process. It is possible to verify that this process is consistent with the ensemble dynamics of the Lindblad master equation. To verify that, it is sufficient to define the projector $\pi(t) = |\psi(t)\rangle \langle\psi(t)|$, compute its stochastic increment by differentiating it as

$$d\pi(t) = d|\psi(t)\rangle \langle\psi(t)| + |\psi(t)\rangle d\langle\psi(t)| + d|\psi(t)\rangle d\langle\psi(t)|, \tag{2.46}$$

substituting Eq. (2.45) and retaining terms up to first order in dt. Finally, when taking the ensemble average of this process, which is equivalent to taking the stochastic average over the processes $dr(t)$, one will find that

$$\frac{\mathbb{E}[d\pi(t)]}{dt} = \mathcal{L}\hat{\rho}(t) \tag{2.47}$$

Details of the calculation are left in the appendix.

2.2.2 Photon counting trajectories

A traditional setup considered in quantum optics is that of a bosonic system which evolves under the Hamiltonian $\hat{H}$, where experimentalists measure the photons leaking out of the system with rate γ by means of a photodetector. The photodetector can give two outcomes, $r = \{0, 1\}$, corresponding to the outcomes *no photon detected* ($r = 0$) and *photon detected* ($r = 1$). The corresponding measurement operators, which also generate the dynamics of the system, are the usual

$$\begin{cases} \hat{B}_0 = \mathbb{I} - idt \left(\hat{H} - i\frac{\gamma}{2} \hat{a}^\dagger \hat{a} \right) \\ \hat{B}_1 = \sqrt{dt\gamma}\hat{a} \end{cases} \tag{2.48}$$

and once again, the probability to detect a photon ($r = 1$) is

$$p_1(t) = dt\gamma \operatorname{Tr}\left[\hat{a}^\dagger \hat{a} \hat{\rho}(t) \right] = dt\gamma \langle n \rangle (t). \tag{2.49}$$

In the limit of $dt \to 0$, the measurement will return a null result at almost every time, and the system will evolve according to the dynamics generated by $\hat{B}_0$. With probability $p_1(t)$, a photon will leak out of the system and it will be observed (assuming the detector has 100% efficiency).

An example of photon counting trajectories is given in the left panel of Fig. 2.1, where the discrete jumps in the number of photons are clearly visible.

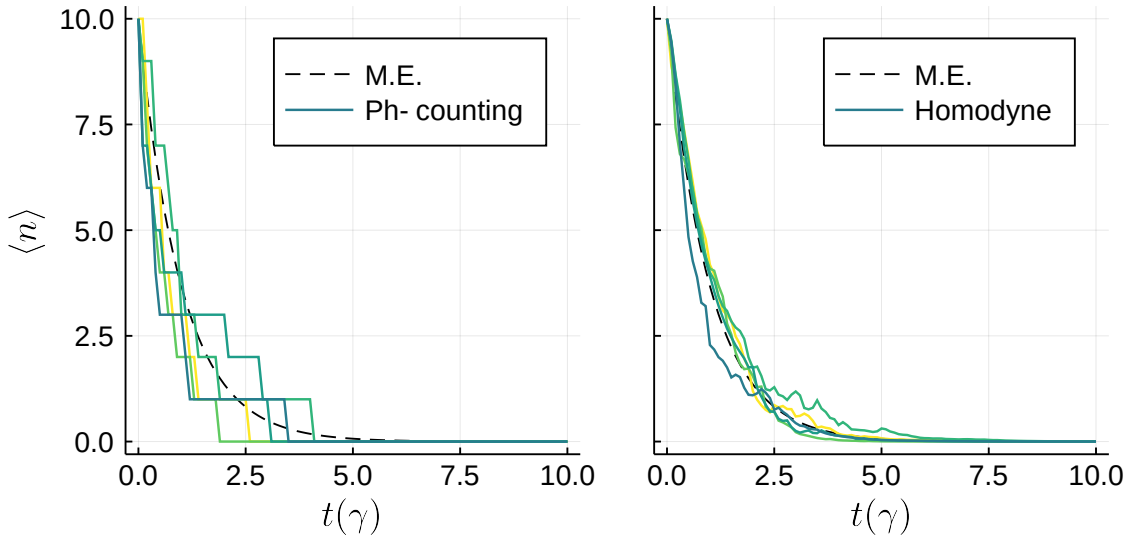

Figure 2.1: Example of different trajectory unravelings of the same Liouvillian. The system is initialized in a fock state with ten photons, and evolves under the action of $\hat{H} = 0$ and jump operator $\hat{L} = \hat{a}$ ($\gamma = 1$). Photon counting (left panel) and Homodyne (right panel) trajectories are plotted with a colored solid line. The exact solution obtained by integrating the master equation is plotted with a dashed black line.

2.2.3 Homodyne trajectories

In the previous paragraph we described the evolution of a lossy optical system through the two operators $\hat{B}_0$ and $\hat{B}_1$ representing a certain continuous measurement being performed on the system. It is possible to verify that the resulting Lindblad master equation is invariant under the transformation

$$\sqrt{\gamma}\hat{a} \to \sqrt{\gamma}\hat{a} + \beta\mathbb{I} \qquad \hat{H} \to \hat{H} - \frac{i}{2}(\beta^\star\hat{a} - \beta\hat{a}^\dagger) \tag{2.50}$$

where β is an arbitrary complex number [21, 82]. If the same transformation is applied to the measurement operators, one obtains

$$\begin{cases} \hat{B}_0 = \mathbb{I} - idt\left(\hat{H} - \frac{i}{2}(\beta^\star\hat{a} + \beta\hat{a}^\dagger) + \frac{\gamma}{2}\hat{a}^\dagger\hat{a}\right) \\ \hat{B}_1 = \sqrt{dt}(\sqrt{\gamma}\hat{a} + \beta\mathbb{I}) \end{cases} \tag{2.51}$$

This shows that the unraveling of the deterministic master equation into stochastic quantum trajectories is not unique. Physically, this means that $\hat{B}_1$ does not measure only the field coming out of the system ($\propto \gamma\langle n\rangle$), but rather the mixing of the output field with a classical field β. This transformation can be realized in an experiment by changing the way we measure the system: instead of using a detector to measure only the outgoing field, one can mix it with the much stronger field of an oscillator. As the quantum system is left untouched and only the measurement setup is modified, it is clear why this should

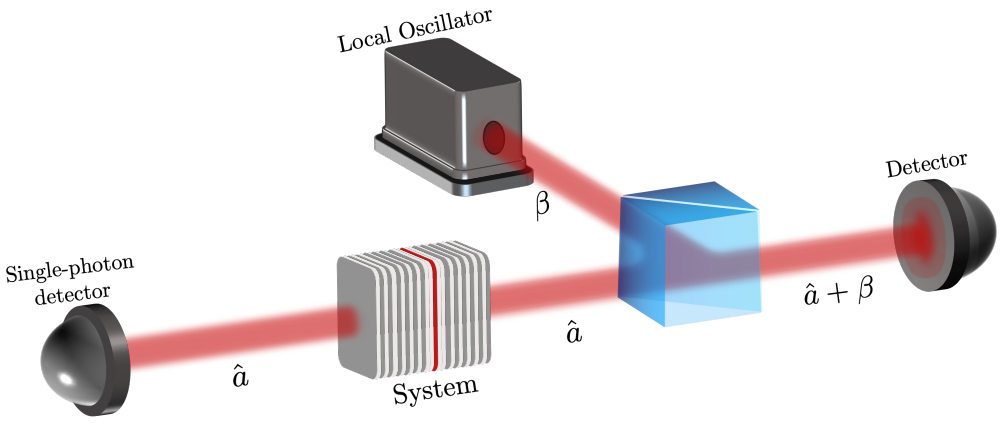

Figure 2.2: Example of experimental setup for photon-counting detection (left), where the output signal is directly measured and homodyne detection (right), where it is mixed with a local oscillator. In practice, in homodyne detection experiments a second detector is used to get the reference value of β, which would be otherwise be subject to big uncertainities.

leave the master equation unchanged. In such a setup, the probability to have a positive measurement is given by

$$p_1 = dt\gamma \langle n \rangle + \sqrt{\gamma}\beta \langle \hat{x} \rangle + |\beta|^2, \tag{2.52}$$

where we have chosen $\gamma \in \mathbb{R}$ and $\hat{x} = \hat{a} + \hat{a}^\dagger$. In the limit when the local field is much stronger than the output field ($\beta/\gamma \to \infty$), the detector will be triggered at every timestep. In fact, to properly compute this limit one should also send $dt \to 0$, as otherwise, due to the strength of the field, one could have more than one detection per timestep (signaled by the fact that $p_1 > 1$, which could no longer be interpreted as a probability) [21, 82]. In this limit, the stochastic nature of the process becomes continuous and one obtains the stochastic differential equation

$$d|\psi(t)\rangle = \left[-idt\left(\hat{\tilde{H}} - i\frac{\hat{L}^\dagger\hat{L}}{2} + \hat{L}\frac{\langle\hat{L}^\dagger + \hat{L}\rangle}{2} - \frac{\langle\hat{L}^\dagger + \hat{L}\rangle^2}{8}\right) + \right.$$
$$\left. + dW(t)\left(\hat{L} - \frac{\langle\hat{L}^\dagger + \hat{L}\rangle}{2}\right)\right]|\psi(t)\rangle\,, \tag{2.53}$$

where $dW(t)$ is a diagonal noise term, which is obtained from the continuous limit of the poissonian process $dr(t)$ previously considered.

Using photon counting or homodyne trajectories to study a system will give the same results once averaged, but the statistical distributions of the trajectories will differ. Most importantly, all observables should give the same average values regardless of the choice of trajectory protocol, even though the fluctuations can differ. An interesting example is given by the informational-theoretic quantity of entropy production: if the system is out of equilibrium, as is usually the case in open systems, it continually produces entropy (also

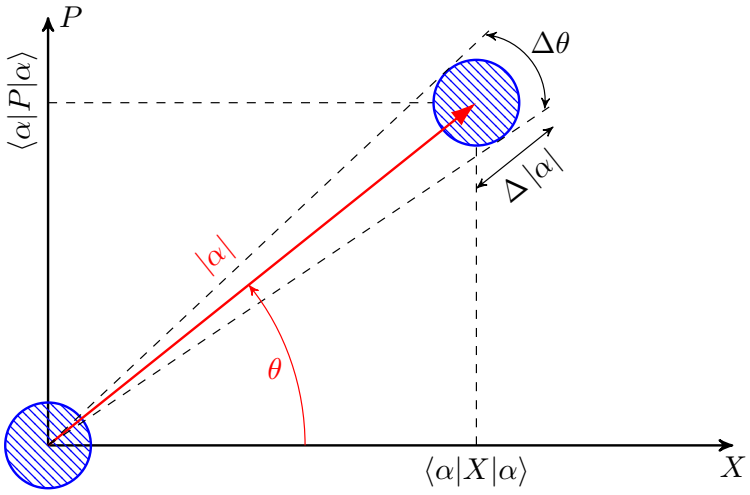

Figure 2.3: Representation of coherent states in the phase space (X, P). The highlighted region corresponds to the region containing the 95% of probability. $|\alpha|$ corresponds to the average number of photons, and θ to the phase. It is clear that in the limit of many photons, the uncertainty on the phase becomes negligible, while the uncertainity on the number of photons remains constant.

at the steady state), and dissipates it into the environment so that the following relation holds

$$\frac{dS[\hat{\rho}]}{dt} = \Phi + \Pi, \tag{2.54}$$

where Π is the rate of entropy production and Φ is the rate of entropy dissipation [99, 100]. It has been shown that those quantities depend on the relative strength of the local oscillator β [101], which is consistent with the fact that entropy-related quantities are linked to fluctuation properties of the trajectories and are not observables [102, 103]. While a considerable theoretical effort has been invested in the field of stochastic thermodynamics [104, 105, 106, 107, 108], there only exist few experimental implementations [109].

2.3 Phase-space representation of bosonic systems

Describing quantum mechanical systems with kets and operators, objects which normally do not commute, causes several complications when performing analytical calculations. An useful set of tools to simplify the description of those systems to a manageable level are phase-space methods, which describe quantum systems through $\mathbb{C}-$numbers. In this section we will describe how the density matrix can be represented as a quasi-probability distribution. Those will be used to construct a phase-space version of the lindblad master equation.

2.3.1 Coherent states

When working with many-body quantum systems, a natural basis is the Fock basis. The Fock basis labels the space according to the number of fermionic or bosonic excitations in each mode, and has the likable property of being orthonormal. Fock states are defined as the quantized eigenstates of the quantum harmonic oscillator, therefore they are naturally quantum objects. A different choice of basis are coherent states, originally introduced by E. Schrödinger when studying the lowest uncertainity eigenstates of the harmonic oscillator. As we will show below, they have a very well defined interpretation in terms of classical objects, and therefore complement Fock states.

It is well known that uncertainity of two observables $\hat{A}$ and $\hat{B}$ in quantum mechanics must satisfy

$$\langle \Delta A \rangle^2 \langle \Delta B \rangle^2 \geq \frac{1}{4} \langle i[\hat{A}, \hat{B}] \rangle^2 , \tag{2.55}$$

where $\Delta\hat{A} = \hat{A} - \langle \hat{A} \rangle$. If we consider the quadrature operators from quantum optics

$$\hat{X} = \frac{\hat{a}^\dagger + \hat{a}}{2}, \qquad \hat{P} = \frac{\hat{a}^\dagger - \hat{a}}{2i}, \tag{2.56}$$

it is possible to show that the state $|\alpha\rangle$ which ensures the equality in Eq. (2.55) must satisfy the equation

$$\Delta\hat{X} |\alpha\rangle = -i\Delta\hat{P} |\alpha\rangle , \tag{2.57}$$

which is verified when

$$(\hat{X} + i\hat{P}) |\alpha\rangle = (\langle \hat{X} \rangle + i \langle \hat{P} \rangle) |\alpha\rangle \tag{2.58}$$

where the coherent states are defined as

$$\hat{a} |\alpha\rangle = \alpha |\alpha\rangle . \tag{2.59}$$

Coherent states are therefore defined as the eigenstates of the destruction operator and minimize the uncertainty. It is therefore easy to see from Fig. 2.3 that, as the number of photons is increased (the distance from the origin in the X, P plane is increased) the uncertainty in the phase becomes negligible. Coherent states can be obtained by displacing the vacuum state $|0\rangle$ by the complex vector α, namely

$$|\alpha\rangle = \hat{D}(\alpha) |0\rangle = \exp\left[\alpha\hat{a}^\dagger - \alpha^\star\hat{a}\right] |0\rangle , \tag{2.60}$$

therefore they all have the same uncertainty as the vacuum (which classically has no uncertainty). For that reason one might interpret the quantum noise of a coherent state as being due only to quantum fluctuations of the vacuum.

2.3.2 Phase-space representations of the density matrix

The expression for the statistical average of measurements of a quantum observable $\hat{A}$ performed on a system with state $\hat{\rho}$ is given by the trace of their product:

$$\langle \hat{A} \rangle = \mathrm{Tr}\left[\hat{\rho}\hat{A}\right]. \tag{2.61}$$

Cahill and Glauber proposed a formalism where, instead of operatorial equations such as the one above, the statistical average is computed with an integral over a distribution space (phase-space) [110], according to standard probability theory :

$$\langle \hat{A} \rangle \rightarrow \langle A \rangle = \int W_s(\alpha, \alpha^\star) A_s(\alpha, \alpha^\star) d\alpha d\alpha^\star \tag{2.62}$$

where $A_s(\alpha, \alpha^\star)$ is a phase-space analogue of the operator $\hat{A}$, $s = \{\pm 1, 0\}$ is an integer labelling one of several phase-space representations. $W_s(\alpha, \alpha^\star)$ provides the integration measure in the $\mathbb{C}^2$ space yielding expectation values, as a probability density function normalized according to

$$\int W_s(\alpha, \alpha^\star) d\alpha d\alpha^\star = 1. \tag{2.63}$$

However, W_s is not a canonical probability distribution as it is possible to verify that the various states $(\alpha, \alpha^\star)$ do not represent mutually exclusive (orthogonal) states, as required by probability theory. Moreover, in general it can take negative values as $W_s(\alpha, \alpha^\star) < 0$. But since W_s has similar mathematical properties of a probability density function, it is called a quasi-probability distribution.

The quasi-probability distribution W_s is defined to be the complex Fourier transform of the characteristic function χ_s according to:

$$W_s(\alpha, \alpha^\star) = \frac{1}{\pi^2} \int \chi_s(\xi, \xi^\star) e^{\alpha \xi^\star - \alpha^\star \xi} d\xi d\xi^\star. \tag{2.64}$$

The characteristic function is defined by

$$\chi_s(\xi, \xi^\star) = \mathrm{Tr}\Big[\hat{\rho} \hat{D}_s(\xi, \xi^\star)\Big], \tag{2.65}$$

where $\hat{D}_s$ is the normally ordered ($s = 1$) anti-normally ordered ($s = -1$) or symmetrically ordered ($s = 0$) exponential

$$\hat{D}_s(\xi, \xi^\star) = \exp\left[\xi \hat{a}^\dagger - \xi^\star \hat{a} + \frac{1}{2} s |\alpha|^2\right]. \tag{2.66}$$

In general, the phase-space analogue of the expectation value of an operator can be deduced from the property (which follows from the definition of $\hat{D}_s$ above),

$$\{(\hat{a}^\dagger)^n (\hat{a})^m\}_s = \frac{\partial^n}{\partial \xi^n} \left(\frac{\partial}{\partial \xi^\star}\right)^m \hat{D}_s(\xi, \xi^\star)|_{\xi^\star = \xi = 0}. \tag{2.67}$$

Expectation values of s-ordered operators can be computed according to

$$\langle \{(\hat{a}^\dagger)^n (\hat{a})^m\}_s \rangle = \int (\alpha^\star)^n \alpha^m \; W_s(\alpha, \alpha^\star) \; d\alpha d\alpha^\star. \tag{2.68}$$

The different orderings give rise to quasi-probability distributions having different likeable properties. Most notably, if $s = -1$ the distribution essentially given by the diagonal elements of the density matrix in the coherent basis,

$$Q(\alpha, \alpha^\star) \equiv W_{-1}(\alpha, \alpha^\star) = \frac{1}{\pi} \langle \alpha | \hat{\rho} | \alpha \rangle \,. \tag{2.69}$$

This is called Housimi Q distribution, and it is useful to perform some analytical calculations and can often be directly measured in experiments.

The value $s = 1$ gives rise to the Glauber-Sudarshan P representation such that $P(\alpha, \alpha^\star) = W_1(\alpha, \alpha^\star)$, where

$$\hat{\rho} = \int P(\alpha, \alpha^\star) d\alpha d\alpha^\star. \tag{2.70}$$

The ensemble generated by the P distribution is a diagonal collection of coherent states. Therefore it is easy to see that the P distribution will be positive in the case of superpositions of classical coherent states. Moreover, as the P representation generates normal-ordered operator averages, it is often used to perform analytical calculations [111, 95].

Lastly, the values $s = 0$, which corresponds to symmetric ordering, generates the Wigner representation $W(\alpha, \alpha^\star) \equiv W_0(\alpha, \alpha^\star)$. The Wigner quasi-probability can be used to efficiently build a semiclassical approximation of the lindblad master equation which accounts for vacuum fluctuations, as we will see in Section 3.4.3.

2.3.3 Phase-space equations of motion

Overall, one wants to compute the equation of motion for the quasi-probability distribution that are equivalent to the Lindblad master equation. After a lengthy calculation it is possible to derive simple conversion rules, where one has to substitute to the operators in the equation of motion of $\hat{\rho}$ particoular Wirtinger derivatives in phase-space [112, 113].

The conversion table for bosonic ladder operators, in the same generic formalism of the references above, is reported here:

$$\hat{\rho} \to W_s(\alpha, \alpha^\star) \tag{2.71}$$

$$\begin{aligned}
\hat{a}\hat{\rho} &\to \left[\alpha - \frac{s-1}{2}\frac{\partial}{\partial \alpha^\star}\right] W(\alpha, \alpha^\star), & \hat{\rho}\hat{a} &\to \left[\alpha - \frac{s+1}{2}\frac{\partial}{\partial \alpha^\star}\right] W(\alpha, \alpha^\star), \\
\hat{a}^\dagger\hat{\rho} &\to \left[\alpha^\star - \frac{s+1}{2}\frac{\partial}{\partial \alpha}\right] W(\alpha, \alpha^\star), & \hat{\rho}\hat{a}^\dagger &\to \left[\alpha^\star - \frac{s-1}{2}\frac{\partial}{\partial \alpha}\right] W(\alpha, \alpha^\star),
\end{aligned}$$

Performing the mapping from operatorial-space to phase-space will yield a partial-differential equation (PDE) in $(\alpha,\ \alpha^\star)$ as opposed to an operatorial equation. The resulting PDE will have the structure of a generalised Fokker-Planck equation [85] and can be studied by solving repeatedly the associated stochastic differential equation [112, 114, 115, 116].

2.3.4 Example: the harmonic oscillator

For example, if we consider a coherently driven harmonic resonator with single-particle losses, described by the Hamiltonian

$$\hat{H} = -\Delta \hat{a}^\dagger \hat{a} - F(\hat{a}^\dagger + \hat{a}), \tag{2.72}$$

and loss operator $\hat{L} = \sqrt{\gamma}\hat{a}$, the resulting phase-space equation of motion will be

$$\frac{\partial W_s(\alpha, \alpha^\star, t)}{\partial t} = L_s W_s(\alpha, \alpha^\star, t), \tag{2.73}$$

where the linear operator L_s is given by

$$L_s = -\frac{\partial}{\partial \alpha}\left[-\gamma\alpha \ + i\Delta\alpha \ + iF\right] + \tag{2.74}$$

$$-\frac{\partial}{\partial \alpha^\star}\left[-\gamma\alpha^\star - i\Delta\alpha^\star - iF\right] + \tag{2.75}$$

$$+\gamma(1-s)\frac{\partial^2}{\partial\alpha\partial\alpha^\star}. \tag{2.76}$$

It is worth noting that L_s identifies a Fokker Planck equation with diffusion coefficient $\gamma(1-s)$ and drift vector $i(\Delta\alpha + iF) - \gamma\alpha$.

As we had discussed in Section 2.3.1, the statistical uncertainity associated with coherent states emerges from quantum fluctuations of the vacuum. In the fokker planck equation those fluctuations arise due to the diffusion term. As the Fokker-Planck equation 2.73 has a positive-definite diffusion term, if the system is initialized so that $W_s(\alpha, \alpha^\star) > 0$ everywhere, it is possible to interpret W_s as a probability distribution at all times.

CHAPTER 3

Numerical methods for open quantum systems

The simulation of Open Quantum Systems on a classical computer is more demanding and complicated than simulating an Hamiltonian problem, as the number of elements in the density matrix is the square of the Hilbert space dimension. Moreover, the dynamical equation, the so-called master equation, being non hermitian is more complicated to treat numerically. In recent years, considerable effort has been invested in developing methods to simulate systems under generic non-equilibrium conditions, such as renormalisation group calculations exploiting the Keldysh formalism [57, 58], dynamical mean field [117], truncated correlation hierarchies [118], corner-space renormalization [69], permutation-invariant solvers [60], full configuration-interaction Monte Carlo [61], cluster factorisations [62] or expansions [63] and tensor-network techniques [65, 67, 68, 66].

Alternatively, exploiting the continuous monitoring interpretation of open quantum systems, several other methods that can be applied to quantum trajectories have been proposed, such as those based on matrix product states [119], cluster ansätzes [120], Gützwiller mean-field [121, 122], semiclassical approaches [1, 123], or with an analytical ansatz [124]. There also exist some variational-inspired methods [64, 125], which have been recently combined with neural networks [3, 73, 74, 75].

In this chapter we will present a brief review of some numerical methods that are routinely used to simulate *finite-size* open quantum systems. We will start with a few details on how to best treat the full master equation in Section 3.1. We will then discuss methods targeting the full steady-state density matrix in Section 3.2 and in Section 3.3 we will give an outline of the developments of tensor-network based techniques for open systems. Variational methods, while relevant, will not be discussed, as the entirety of Chapter 5 will be devoted to the topic. Section 3.4 will present a discussion of stochastic trajectory methods, both exact and approximate. To conclude, in Section 3.5 we will address how to treat disordered systems, in particular by introducing in Section 3.5.3 the

optimal stochastic unraveling of disordered open quantum systems [2].

Contents

3.1 The Lindblad master equation

As discussed in the previous chapter, the Lindblad master equation

$$\frac{d\hat{\rho}_S}{dt} = -i\Big[\hat{H}_S, \hat{\rho}_S\Big] + \sum_i \left(\hat{L}_i \hat{\rho}_S \hat{L}_i^\dagger - \frac{1}{2}\Big\{\hat{\rho}_S, \hat{L}_i^\dagger \hat{L}_i\Big\}\right), \tag{2.27}$$

encodes the dynamics of an open quantum system. This equation identifies a system of Ordinary Differential Equations (ODE) of the form $\dot{\hat{\rho}} = \mathcal{L}\hat{\rho}$, which can be numerically integrated to obtain the time-evolution of $\hat{\rho}(t)$. If a metastable state arises in the quantum system, the Jacobian of the ODE might become ill-conditioned for a short time. This phenomenon (which is otherwise not very well mathematically defined) is known as stiffness, and in general it has the consequence that explicit solvers (such as those based on the Taylor expansion of the ODE) will introduce non-negligible errors or require exponentially small timesteps; implicit solvers, which are more computationally demanding, circumvent the problem [126]. Methods based on the Runge-Kutta of $4^{\text{th}}(5)$ order with adaptive step size [127] are in general the best for reliability and efficiency. It should be noted that textbook versions of Runge-Kutta 4(5) perform well only on non-stiff problems, and the

Dormann-Prince variant [128] should be preferred, as it can handle semi-stiff equations[1]. If the system is known for certain to be non-stiff, the *TSIT5* solver [129] gives the best performance.

For small systems a viable strategy is to compute all the elements of $\mathcal{L}$ and store it as a sparse matrix according to Eq. (A.4). If l is the local Hilbert space dimension, N is the number of local spaces and s a constant, the computational complexity of building the full Liouvillian is roughly[2] $\mathcal{O}(l^{2(1+N)}s^{2(1+N)})$. On a desktop computer it is typically feasible to construct the Liouvillian of a spin-1/2 system only up to 9 lattice sites ($\dim(\mathcal{H}) \approx 500$). It is therefore intractable to build this matrix already for relatively small lattices, as the cost of constructing the ODE would exceed by far the time taken to solve it. Fortunately, efficient ODE solvers do not need to work with a full matrix, but only need a way to evaluate the action of Eq. (2.27) on an arbitrary input[3]. This way, it is possible to solve systems up to approximately 16 spins ($\dim(\mathcal{H}) \approx 6 \cdot 10^4$).

3.1.1 Factorized methods

The time-integration of the Lindblad master equation (2.27) for extended lattice systems is a formidable task. The exponential increase of the Hilbert space makes a direct integration of the master equation unfeasible already for relatively small system sizes.

A possible approximated approach for this problem is given by mean field methods, such as those based on Gützwiller's factorised ansatz for the density matrix [130]. Such approach can be generalised to clusters (or plaquettes) of sites [62]. The general idea is that even if it is not possible to integrate the master equation $d\hat{\rho}/dt = \mathcal{L}\hat{\rho}$ for the whole system, it is possible to do so for a smaller partition. Consider the subdivision of all the sites $[1...N]$ of the system into $N_\mathcal{C}$ disjoint clusters, each with a number $[1, N[$ of sites, so that the union of all clusters reconstructs the system. This subdivision can be used to construct a factorised cluster ansatz for the global density matrix

$$\hat{\rho} \approx \otimes_{i=1}^{N_\mathcal{C}} \hat{\rho}_{\mathcal{C}_i}, \tag{3.1}$$

where $\hat{\rho}_{\mathcal{C}_i}$ is the density matrix of the $i-$th cluster (which could be as small as a single site). Factorised methods keep track of quantum correlations within the cluster, but withhold only classical correlations among clusters. Assuming that $\mathcal{L}$ is composed only of local and

[1] The Runge-Kutta-Dormann-Prince from [128], known as DP5 or DOPRI is the default method used by *ode45* in Matlab and SciPy. In Julia's *DifferentialEquations.jl* package it is implemented under the name DP5.

[2] The complexity is derived assuming operators are stored as sparse matrices. Dense matrices would lead to worse asymptotical performance. The complexity of taking the tensor product of two $l \times l$ sparse matrices with s nonzero entries per row is $\mathcal{O}(l^2 s^2)$. The chain of tensor products needed to build local operators require asymptotically $\mathcal{O}(l^{1+N}s^{1+N})$ many operations, and the most-complicated term is the one of the form $L \otimes L$ which has squared this complexity.

[3] Using a linear map instead of a matrix is a fundamental trick in numerical computation. It is important to note that, contrary to solving matricial ODEs, solving ODEs with linear maps requires extra care in interpreted languages such as python, otherwise performance might quickly degrade. In particular, if the user-provided map is not compiled, performance will quickly degrade. Compilation is typically ensured by using Cython/Numba in Python. Compiled languages such as Fortran/C++/Julia do not need to worry.

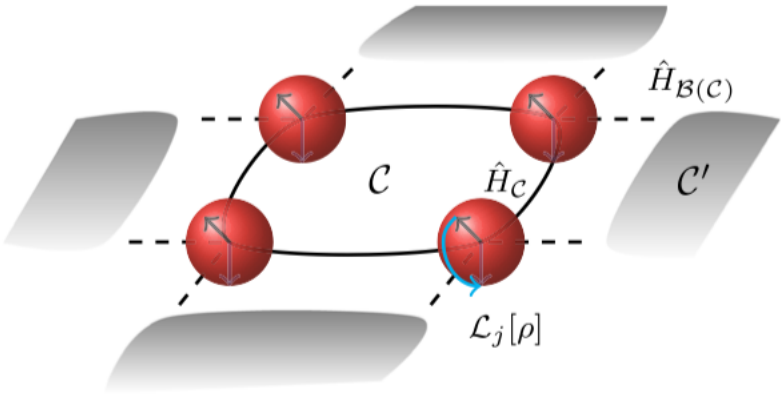

Figure 3.1: Sketch of the cluster mean field approach in a dissipative system of interacting spin-1/2 particles. The figure refers to a 2×2 cluster on a two dimensional square lattice. Figure from [62].

2-local operators[4], it can be casted into the form,

$$\mathcal{L} = \sum_i \mathcal{L}_i + \sum_{i,j} \mathcal{L}_{i,j}, \tag{3.2}$$

where $\mathcal{L}_i$ is the local Liouvillian term for the $i-$th cluster, given by the local Hamiltonian and dissipation, and $\mathcal{L}_{i,j}$ is the coupling term between clusters i and j, given by coherent hopping or non-local dissipators. Substituting the ansatz of Eq. (3.1) and the factorised Liouvillian into Eq. (2.27) gives the reduced master equation for every cluster $\mathcal{C}_i$,

$$\frac{d\hat{\rho}_{\mathcal{C}_i}}{dt} = \mathcal{L}_i \hat{\rho}_{\mathcal{C}_i} + \sum_{\langle i,j \rangle} \mathcal{L}_{i,j} \hat{\rho}_{\mathcal{C}_i} \hat{\rho}_{\mathcal{C}_j}. \tag{3.3}$$

This is the composed by the local term $d\hat{\rho}_{\mathcal{C}_i}/dt = \mathcal{L}_i \hat{\rho}_{\mathcal{C}_i}$ plus a coupling to the neighbouring clusters. The coupling term can be computed self-consistently at every timestep with a computational cost comparable to that of computing a local observable because the ansatz is factorised, leading to the self-consistent master equation

$$\frac{d\hat{\rho}_{\mathcal{C}_i}}{dt} = (\mathcal{L}_i + \sum_{\langle i,j \rangle} \mathcal{L}_i^{[\mathrm{j}]}) \hat{\rho}_{\mathcal{C}_i}, \tag{3.4}$$

$$\mathcal{L}_i^{[\mathrm{j}]} = \mathrm{Tr}\big[\mathcal{L}_{i,j} \hat{\rho}_{\mathcal{C}_j}\big], \tag{3.5}$$

Therefore, with this approximation the computational complexity of time-evolving the system is no longer proportional to $\mathcal{O}(l^{2N})$ but rather to $\mathcal{O}(N_{\mathcal{C}} l_{\mathcal{C}}^2)$. For the Gützwiller factorisation (clusters of size 1) this further simplifies to $mO(Nl^2)$. Furthermore, if one assumes that all clusters are equivalent, the equation simplifies further and the multiplicative term N disappears. Even if in this formulation we used the factorisation of the Hilbert

[4]However the method can be easily generalised to any K-local super-operator

space to simplify the full master equation, it has recently been proposed to combine the factorised wavefunction ansatz together with photon-counting [121, 120] and homodyne [122] trajectories.

3.1.2 Symmetries and permutation-invariant solvers

In general, one must always choose a basis of the Hilbert space to express the various operators and the state. An open quantum system is said to be invariant under the action of some symmetry operator $\hat{J}$ if

$$\left[\hat{J}, \hat{H}\right] = \left[\hat{J}, \hat{L}_i\right] = 0 \quad \forall i, \tag{3.6}$$

that is, if the generator of the symmetry commutes with the Hamiltonian and all the loss operators [37]. Symmetries reduce the effective size of the Hilbert space if a suitable basis is chosen. The dynamics of the Liouvillian will also factorise into sectors that do not talk to each other, further reducing the memory requirements.

An example of an interesting symmetry on open quantum systems is the quadratically driven, quadratically dissipating Kerr model [53, 131, 132]. When dealing with periodic homogeneous lattice systems it might be interesting to consider the set of translations, point rotations and reflections. In an N-dimensional lattice of size $L \times L \times ... \times L$, there are a total of $\mathcal{O}(L^N)$ symmetries, which would reduce polynomially the total size of the Hilbert space.

Remarkably, homogeneous systems with long-range interaction such as the Dicke are invariant under permutations of the modes. As there exist $N!$ permutations of N variables, this is a very strong symmetry that, when implemented, provides an exponential speed-up of the computations [60].

3.2 Methods determining the steady-state

Computing the full time-evolution of a dissipative system, either through the ODE or by using trajectory-based methods (see Section 3.4) can be computationally demanding. Often when studying open systems we are only interested in the properties of the steady state $\hat{\rho}_{ss}$, which is the 0-eigenvalue, defined by the equation

$$\mathcal{L}\hat{\rho}_{ss} = 0. \tag{3.7}$$

There exist a few techniques that target directly the steady state without computing the whole time-dynamics.

3.2.1 Direct solution of the linear system

Equation (3.7) identifies a system of linear equations that can be solved rather efficiently by means of linear solvers, developed extensively in mathematical optimisation theory. Solving this equation for the steady state is usually much simpler than integrating the temporal dynamics of the quantum system. Thus, it can be done faster and for larger systems.

The main issue with this approach is the fact that Eq. (3.7) does not have a zero-dimensional kernel, but all matrices $\hat{M} = t\hat{\rho}_{ss}$ with $\mathrm{Tr}\left[\hat{M}\right] = t$ are valid solutions. Savona et al. proposed originally in [132] to add as an additional constraint the trace condition $\mathrm{Tr}\left[\hat{M}\right] = 1$, by either adding another equation or substituting a line of the Liouvillian. The resulting system can then be solved with a linear-algebra solver. As the system is not hermitian, we found that the best performance is achieved by using the BiConjugate-Stabilized Gradient method [133, 134], but it is possible that future research on linear algebra solvers might find better solvers. We also remark that those solvers can be executed efficiently on a GPU, with a notable speedup (typically an improvement of two orders of magnitude) compared to a standard CPU implementation.

3.3 Tensor networks methods

Tensor Network methods have had a remarkable success in modern numerical physics because of their effectiveness in representing quantum states with finite computational resources. The problems those method try to address is to encode with non exponential resources the wavefunction

$$|\Psi\rangle = \Psi_{\sigma_1,\cdots,\sigma_N} |\sigma_1, \cdots, \sigma_N\rangle \,, \tag{3.8}$$

where $\sigma_1, \cdots, \sigma_N \in \mathcal{H}_1 \otimes \ldots \mathcal{H}_N = \mathcal{H}$.

The fundamental idea of tensor networks consist in approximating the many-body wavefunction with a product of many contracted tensors. The so-called *Matrix Product State* (MPS) ansatz to express the wavefunction of a 1D system with N sites is such an example:

$$\Psi_{\sigma_1,\cdots,\sigma_N} = \sum_{i_1,i_2,\ldots,i_N} A^{i_1,i_2}_{\sigma_1} A^{i_2,i_3}_{\sigma_2} \ldots A^{i_N,i_1}_{\sigma_N}, \tag{3.9}$$

where the N local tensors $A^{i,j}_{\sigma_i}$ have 3 dimensions: σ_i labels the elements of the local Fock space, and i, j are fictitious indices used to introduce correlations among sites. We remark that if i, j where absent, this would correspond to the Gützwiller factorised ansatz with no correlations introduced in Section 3.1.1. The amount of correlations that can be encoded in such an ansatz is proportional to the bond-dimension M. This is the size of the space spanned by those unphysical indices, and it determines the total dimension of a MPS ansatz, which is $l\ M^2$, where l is the local Hilbert space dimension.

When applying an operator to the wavefunction $\Psi(\sigma_1, \cdots, \sigma_N)$, the physical indices $\sigma_1, \cdots, \sigma_N$ are summed together with those of the operator. The MPS ansatz is extremely effective in $1D$ because it is possible to compute expectation values $\langle\Psi|\hat{H}|\Psi\rangle$ by performing those contractions exactly with a numerically efficient algorithm. For $D \geq 2$, however, the sum over all contractions is known to be an hard problem and it cannot be treated exactly. Extensive research is still ongoing to find efficient yet general schemes to compute those contractions in 2D.

3.3.1 Matrix Product Density Operator

For open-quantum systems one must, however, approximate the density matrix and potentially ensure its positive-definite character, a computationally hard problem. In the following we will consider the density matrix written in the form,

$$\hat{\rho} = \sum_{\sigma_1,\tilde{\sigma}_1\ldots,\sigma_N,\tilde{\sigma}_N} \rho_{\sigma_1,\tilde{\sigma}_1,\ldots,\sigma_N,\tilde{\sigma}_N} \left|\sigma_1,\ldots,\sigma_N\right\rangle \left\langle\tilde{\sigma}_1,\ldots,\tilde{\sigma}_N\right|, \tag{3.10}$$

where $(\sigma_1, \tilde{\sigma}_1) \in \mathcal{H} \otimes \mathcal{H}$ are the two indices that label the local space of the density matrix. In the past years, several schemes have been proposed to approximate the function $\rho_{\sigma_1,\tilde{\sigma}_1,\ldots,\sigma_N,\tilde{\sigma}_N}$. This approach suffers from the same limitations as product states when used in more than 1 dimension in addition to problems related to enforcing the positivity of the state. The simplest ansatz is based on the Choi's isomorphism [86] which associates to a density matrix a pure state in an Hilbert space of doubled dimension. The vectorised density matrix is described with a matrix product state, which can be easily done by performing the replacement

$$A_{\sigma_l}^{i_l,i_{l+1}} \to A_{\sigma_l,\tilde{\sigma}_l}^{i_l,i_{l+1}}, \tag{3.11}$$

and maintaining the same contraction structure of Eq. (3.9), valid in the 1D case. In doing so, one obtains the form of the *Matrix Product Operator* (MPO) given by [135],

$$\rho_{\sigma_1,\tilde{\sigma}_1,\ldots,\sigma_N,\tilde{\sigma}_N} = A_{\sigma_1,\tilde{\sigma}_1}^{i_1,i_2} A_{\sigma_2,\tilde{\sigma}_2}^{i_2,i_3} \ldots A_{\sigma_N,\tilde{\sigma}_N}^{i_N,i_1}, \tag{3.12}$$

where we use the Einstein's summation convention, implying that contracted indices are summed upon. This firm is equivalent to the MPS ansatz where every local tensor has two physical dimensions $(\sigma_i\tilde{\sigma}_i)$ instead of one (σ_i). While this ansatz has proven effective and very simple, it does not guarantee that the resulting matrix is positive definite, and as such cannot always guarantee physical results, especially when combined with time-evolutions along very long time scales. Moreover, checking if a matrix is positive definite is an NP-hard problem and therefore cannot be done when systems become relatively large [136].

Time evolution

Since the matrix product operator ansatz represents mixed states as pure states in a doubled Hilbert space, the time-evolution can be performed by interpreting the Liouvillian super-operator $\mathcal{L}$ as a matrix, and using the traditional time-evolving-block-decimation (TEBD) algorithm [137]. While it is not hard to initialise the ansatz in a valid (positive) state, performing a Trotter-approximation of the time-evolution will introduce errors in the eigenvalues of the density matrix, which is not guaranteed to remain positive definite along the evolution. As such, it is not possible to integrate the MPO master equation for arbitrarily long time intervals. Regardless, several interesting results have been obtained for dissipative systems with this technique, which has recently been generalised to 2D systems [66].

Alternatively, it is possible to perform a variational optimisation of the MPO ansatz targeting the steady state, in a spirit similar to what will be described in Chapter 5. For example, Cui et Al. have proposed a method for finding the steady-state by searching

for the ground state of the *effective* Hamiltonian $\mathcal{L}^\dagger\mathcal{L}$ [65], while Savona et al. improved on this result by only searching for the null eigenstate of a local $\mathcal{L}_i$, and sweeping along all sites [64]. A limit of both techniques is that the computational cost scales with the inverse of the squared Liouvillian gap, contrary to real-time evolution which will decay to the steady state with an exponential rate equal to the gap.

3.3.2 Locally purified tensor networks

An improvement to this last ansatz is achieved by the *Locally Purified Density Operator* (LPDO), constructed through a purification procedure, which ensures positive states. The underlying idea is that of the purification ansatz. It is possible to build a valid mixed state by taking the reduced trace of a pure state in a bigger Hilbert space, similar to the construction used in Section 2.1.3. There is freedom in the way one defines the extended space, and it is unclear if one definition is better than others. In general, the extended space is built by coupling to each local site an ancilla of dimension $\tilde{l}$, and then building the MPS wavefunction of this extended system $\mathcal{H}\otimes\mathcal{A} = \mathcal{H}_1\otimes\mathcal{A}_1\otimes\dots\mathcal{H}_N\mathcal{A}_N$,

$$\Psi_{\sigma_1,a_1,\dots,\sigma_N,a_N} = A^{i_1,i_2}_{\sigma_1,a_1}\dots A^{i_N,i_1}_{\sigma_N,a_N}. \tag{3.13}$$

The purification ansatz is built by tracing out the ancillary degrees of freedom of $|\Psi\rangle\langle\Psi|$, obtaining the *reduced* density matrix of the physical system

$$\rho_{\sigma_1,\tilde{\sigma}_1,\dots,\sigma_N,\tilde{\sigma}_N} = \sum_{a_1,\dots,a_N} A^{i_1,i_2}_{\sigma_1,a_1}A^{\tilde{i}_1,\tilde{i}_2}_{\tilde{\sigma}_1,a_1}\dots A^{i_N,i_1}_{\sigma_N,a_N}A^{\tilde{i}_N,\tilde{i}_1}_{\tilde{\sigma}_N,a_N} \tag{3.14}$$

This ansatz is built to be positive-definite, as it has the form $AA^\dagger$ (as well as every local term $A^{i_k,i_{k+1}}_{\sigma_k,a_k}A^{\tilde{i}_k,\tilde{i}_{k+1}}_{\tilde{\sigma}_k,a_k}$ is locally positive definite). We remark that the locally-purified MPO and the vectorised MPO are not equivalent. In fact, while the first might seem more refined as it enforces some likeable physical property, it has been shown that the bond dimension of a LPDO is not bounded by that of a vectorised MPDO [138, 139]. This signals that LPDO are not as efficient encoders as MPDOs. However, when one needs to compute information-theoretical quantities such as fidelity or mutual information, even small imaginary or negative eigenvalues of the density matrix introduce a significant error.

Time-evolution

This ansatz was originally introduced to describe mixed thermal states, which evolve under the action of an Hamiltonian (or an imaginary Hamiltonian). In fact, it is relatively simple to evolve those states while preserving their internal structure under the action of an hermitian operator, by acting first to the left, then to the right side of the density matrix. Unfortunately, the Liouvillian operator has a mixed term $\hat{L}\hat{\rho}\hat{L}^\dagger$, which cannot be easily treated. Therefore, to maintain the purified structure one must resort to elaborate schemes, which have proven to be computationally very costly [68].

We would like to conclude by remarking remark that tensor networks have shown to be relevant not only for encoding quantum states, but also for general machine learning problems [77, 140, 141].

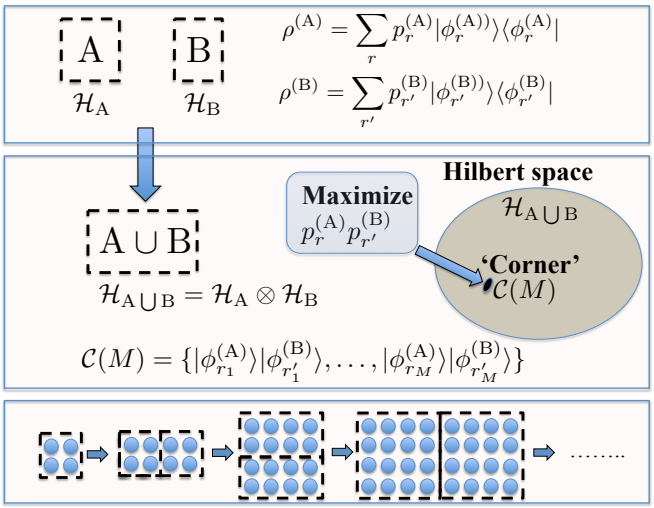

Figure 3.2: Sketch of the corner space renormalisation method. In the top panel the density matrices of the two disjoint systems A and B are shown. In the middle panel the merged system $A \cup B$ and the corner space are sketched. From Ref.[69].

3.3.3 Corner space renormalisation

One of the main problems encountered when simulating large many-body systems is the exponential growth of the Hilbert space, and therefore of the computational cost. The Corner-space renormalisation method is a recently developed technique to determine the steady-state of an open-quantum system, targeting the relevant subspace of the Hilbert space, the so-called *corner space*, for increasingly large systems. By construction, the method is not suitable for steady states with a large entropy, because the corner space is built by truncating parts of the Hilbert Space with negligible occupation. However, it has been used to study spin systems at criticality [43], frustrated bosonic lattices [142], second-order transitions in bosonic lattices [132] and non-Markovian (incoherently driven) Hubbard lattices [91].

The idea of the algorithm, sketched in Fig. 3.2, is depicted in what follows. Consider two systems A and B with Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$. It is essential to consider systems small enough so that they can be solved exactly (with any of the methods described before) in order to obtain the steady-state density matrices $\hat{\rho}_{ss}^{(A)}$ and $\hat{\rho}_{ss}^{(B)}$. When targeting homogeneous systems, one can take $A = B$. Each density matrix can be diagonalised as

$$\hat{\rho}_{ss}^{(A)} = \sum_r p_r^{(A)} \, |\psi_r^{(A)}\rangle\langle\psi_r^{(A)}| \,, \tag{3.15}$$

$$\hat{\rho}_{ss}^{(B)} = \sum_r p_r^{(B)} \, |\psi_r^{(B)}\rangle\langle\psi_r^{(B)}| \,, \tag{3.16}$$

where the states $|\psi_r^{(A/B)}\rangle$ form an orthonormal basis of the Hilbert space $\mathcal{H}_{A/B}$, and

$p_r^{(A/B)}$ give the corresponding probabilities such that $\sum_r p_r^{(A/B)} = 1$.

The corner method aims at constructing a truncated basis for the merged system $\mathcal{H}_{A\cup B} = \mathcal{H}_A \otimes \mathcal{H}_B$. The corner-space of dimension M, $\mathcal{C}(M) \subset \mathcal{H}_{A\cup B}$ is built by composing the basis elements of $\mathcal{H}_A$ and $\mathcal{H}_B$ according to some rule. The rule originally proposed in [69] takes the M most probable basis elements of the joint system $|\psi_{r_1}^{(A)}\rangle\, |\psi_{r'_1}^{(B)}\rangle$, ranked according to the joint probability $p_r^{(A)} p_{r'}^{(B)}$. In other words, the corner space is spanned by the orthonormal basis

$$\mathcal{C}(M) = \left\{ |\psi_{r_1}^{(A)}\rangle\, |\psi_{r'_1}^{(B)}\rangle\,,\, |\psi_{r_2}^{(A)}\rangle\, |\psi_{r'_2}^{(B)}\rangle\,, \ldots,\, |\psi_{r_M}^{(A)}\rangle\, |\psi_{r'_M}^{(B)}\rangle \right\}, \tag{3.17}$$

where

$$p_{r_1}^{(A)} p_{r'_1}^{(B)} \geq p_{r_2}^{(A)} p_{r'_2}^{(B)} \geq \cdots \geq p_{r_M}^{(A)} p_{r'_M}^{(B)}, \tag{3.18}$$

(where indices can repeat themselves, and r_i and r_j can be identical). For the procedure to be successful, the total probability captured by the $\mathcal{C}(M)$ basis must be

$$\sum_{i=1}^{M} p_{r_i}^{(A)} p_{r'_i}^{(B)} \approx 1 \tag{3.19}$$

If this was not to be the case, truncation errors might arise. This condition can be met by systems with moderate entropy. The precise value of M is determined by increasing it until convergence is reached for a target observable. Once a truncated basis $\mathcal{C}(M)$ is built, the whole procedure can be repeated to build the corner space for increasingly large systems.

The structure of the state constructed by means of such iterative renormalisation method is that of a tree-tensor network [143]. The tree network structure is known to be capable of encoding correlations in arbitrary dimensions, as has been further confirmed by the various study of critical systems with the corner-space method [43, 132, 144, 91].

3.4 Trajectory-based methods

Until now we have only described methods that target the density matrix, either exactly or through controlled approximations. A different approach when dealing with open quantum systems consists in not treating the whole density matrix, but rather approximate it with a finite set of pure quantum states, usually obtained by working with the quantum trajectories introduced in Section 2.2. This is a similar approach to what is done with Fokker-Planck equations, where solving the partial differential equations for the distribution probability is a much harder problem than sampling individual stochastic trajectories [145, 146].

Numerical methods based on the unravelling of the master equation approximate the density matrix $\hat{\rho}(t)$ according to

$$\hat{\rho}(t) \approx \frac{1}{N} \sum_{i=1}^{N} |\psi_i(t)\rangle \langle\psi_i(t)|\,, \tag{3.20}$$

where the set of N wavefunction $\{\psi_i\}$ are sampled stochastically by solving an equal number of times the stochastic Schrödinger equation [111]. If the size of the Hilbert space is sufficiently large, reconstructing the whole density matrix is computationally unfeasible. Quantities of interest such as observables and correlation functions are computed by taking stochastic averages over the ensemble of sampled trajectories according to the formula

$$\langle \hat{O} \rangle \approx \frac{1}{N} \sum_{i=1}^{N} \langle O \rangle_i = \frac{1}{N} \sum_{i=1}^{N} \langle \psi_i(t) | \hat{O} | \psi_i(t) \rangle \,. \tag{3.21}$$

Care should be taken when computing two-time correlation functions, as their expression in terms of stochastic averages requires the use of the quantum regression theorem [21, 82, 97, 147]. From Eq. (3.20) one can easily see that the density matrix under this approximation has at most N independent eigenvalues and states, and, similarly, at most rank N. This is a controlled approximation, as it is possible to increase the number of trajectories N until convergence is reached. Of course the value of N at which convergence will be reached varies depending on how entropic the system is and what particular quantity one is computing.

This approximation is particularly effective when working with almost pure (low Von-Neumann entropy) states, which correspond to low-rank density matrices. In this case, the number of trajectories needed to converge is limited. The method is instead inefficient when dealing with high-entropy states, such as the maximally mixed states which require many ($\mathcal{O} \sim \dim[\mathcal{H}]$) eigenstates [148].

From an algorithmic point of view, wavefunction based methods trade memory for the number of operations: when solving the Lindblad master equation for the time-evolution of the full density matrix, one needs to compute the superoperator-operator product $\mathcal{O}(\frac{t_f - t_i}{dt})$ times, with a memory requirement of $\mathcal{O}(\exp[2D])$, where D is the number of lattice sites in the system. With a wavefunction based method, instead, one evolves N trajectories, therefore requiring $\mathcal{O}(N\frac{t_f - t_i}{dt})$ matrix-vector products, with a memory requirement of $\mathcal{O}(\exp[D])$. While still having exponential memory requirements, the decrease from $\mathcal{O}(\exp[2D])$ to $\mathcal{O}(\exp[D])$ allows researchers to push numerical simulations much farther. Moreover, as each stochastic trajectory $\psi_i(t)$ is an independent stochastic process the algorithm can be easily parallelised. This is opposed to the time-evolution of the density matrix, which requires solving a single ODE, a task that traditionally cannot be scaled to more than a handful of computing cores. However, we would like to point out that a set of ODE solvers known as PITA (parallel in time) [149] or PFASST (parallel full approximation scheme in space and time) [150], which provide decent, though far from ideal, scaling when distributed across compute nodes, have been developed in recent years.

3.4.1 Photon-counting trajectories

We have shown in Section 2.2.1 that it is possible to perform two (actually, several) different unravellings of the master equation, both of which are based upon the stochastic Schrödinger equation. While all the unravellings lead to a stochastic differential equation, they must be treated differently in order to achieve numerical efficiency.

In case of photon-counting, the differential equation derived in Section 2.2.2 takes the

form of

$$\frac{d\,|\psi(t)\rangle}{dt} = \left[i\left(-\hat{\tilde{H}} - \frac{i}{2}\sum_i \langle \hat{L}_i^\dagger \hat{L}_i \rangle_{\psi(t)} \right) + \sum_i dR_i(t) \left(\frac{\hat{L}_i}{\sqrt{\langle \hat{L}_i^\dagger \hat{L}_i \rangle_{\psi(t)}}} - 1 \right) \right] |\psi(t)\rangle \tag{3.22}$$

where $\hat{\tilde{H}} = \hat{H} - \frac{i}{2}\sum_i \hat{L}_i^\dagger \hat{L}_i$ is the effective non-Hermitian Hamiltonian and $\{\hat{L}_i\}$ is the set of jump operators. Molmer, Castin and Dalibard showed in 1992 that this equation can be interpreted as if the system is evolving according to the effective non-hermitian Hamiltonian $\hat{\tilde{H}}$ [151]. This leads to a decrease in the norm of the wavefunction given by $\langle\psi(t+dt)|\psi(t+dt)\rangle = \langle\psi(t)|\psi(t)\rangle - dp(t) + \mathcal{O}(dt^2)$, where we retain only terms linear in dt,

$$dp(t) = dt\sum_i p_i(t) = dt\sum_i \frac{\langle\psi(t)|\hat{L}_i^\dagger \hat{L}_i|\psi(t)\rangle}{\langle\psi(t)|\psi(t)\rangle}, \tag{3.23}$$

and the timestep dt must be chosen so that $dp \ll 1$. The square modulus of states in quantum mechanics are interpreted as probabilities. The fact that $\hat{\tilde{H}}$ does not conserve the total probability means that it does not represent all possible outcomes of the time.evolution In particular, with probability $1 - dp(t)$ the time-evolved state will be given by

$$\psi(t+dt)_{1-dp(t)} = \exp\left[-i\hat{\tilde{H}}dt\right]\psi(t), \tag{3.24}$$

while with probability $dp(t)$ the system has been *measured* by the environment and its state is macroscopically changed by the jump operator $\hat{L}_i$ (where the $i-$th operator is selected with probability $dp_i(t)$). Based on this idea, modern algorithms implemented in QuTip [152, 153] or QuantumOptics.jl [154] employ a faster than the clock algorithm consisting of three steps:

1 Chose a random number $r \in [0,1]$, encoding the probability that a quantum jump occurs;

2 Solve the dynamics generated by $\hat{\tilde{H}}$ for the state $|\psi(t)\rangle$ until the time t' when the norm of the wavefunction satisfies $\langle\psi(t')|\psi(t')\rangle = r$;

3 At time t' select the $i-$th quantum jump with probability $p_i(t')$, and apply the corresponding $\hat{L}_i$ operator to $|\psi(t')\rangle$, and normalise the resulting state, obtaining

$$\left|\psi'(t')\right\rangle = \frac{\hat{L}_i\,|\psi(t')\rangle}{\langle\psi(t')|\hat{L}_i^\dagger \hat{L}_i|\psi(t')\rangle} \tag{3.25}$$

4 Use the state $|\psi'(t')\rangle$ as starting point of the algorithm again, until the end time of the simulation is reached.

Interestingly, by exploiting the fact that jump-diffusion processes are particular examples of Random Ordinary Differential Equations (RODE) [155], in the future it might be possible to introduce more advanced, adaptive methods [156]. Notably, computing the different trajectories in parallel can be done with practically zero overhead, and scaled to big computing clusters [157].

3.4.2 Homodyne trajectories

The other type of quantum trajectory introduced in Section 2.2.3 are homodyne trajectories. In the canonical limit of infinite reference field amplitude, the stochastic Schrödinger equation is rewritten as

$$d\left|\psi(t)\right\rangle = \left[-idt\left(\hat{\tilde{H}} + \frac{i}{2}\left\langle \hat{L}^\dagger + \hat{L}\right\rangle \hat{L}\right) + dW(t)\left(\hat{L} + \frac{1}{2}\left\langle \hat{L}^\dagger + \hat{L}\right\rangle\right)\right]\left|\psi(t)\right\rangle . \tag{3.26}$$

Where, for simplicity, we considered the case of only one jump operator $\hat{L}$. In the equation above $dW(t)$ is a real-valued Wiener process, which generates a $\delta-$correlated noise [158]. The above equation is a stochastic differential equation with drift vector $\mu = \hat{\tilde{H}} + \frac{i}{2}\left\langle \hat{L}^\dagger + \hat{L}\right\rangle \hat{L}$ and diffusion vector $\sigma^2 = \hat{L} + \frac{1}{2}\left\langle \hat{L}^\dagger + \hat{L}\right\rangle$.

A simple, order method to solve this kind of equations is the following (Euler-Mayurma):

- At every timestep t, compute $dW(t)$ by generating a random number drawn from a gaussian distribution with mean 0 and variance dt;
- Use $dW(t)$ to compute the new state according to $\left|\psi(t+dt)\right\rangle = \left|\psi(t)\right\rangle + d\left|\psi(t)\right\rangle$;

While this algorithm is very easy to implement, it only has order $\mathcal{O}(\sqrt{dt})$ and thus requires very small timesteps to converge correctly. More advanced, higher order algorithms are described in the book by Kloeden and Platen [98], but they do not scale well for systems with more than one dissipator. In those cases, more advanced adaptive timestepping algorithms improve the performance by several orders of magnitude [159, 160, 161].

3.4.3 Semiclassical methods

Photon-counting and homodyne trajectories are a powerful tool when describing quantum systems, as they evolve numerically exact trajectories. The error arising from the stochastic sampling of observables can be made arbitrarily small by increasing the number of trajectories. However, since those methods encode the pure states into a vector representation of the full Hilbert space, they are also plagued by an exponentially-increasing computational cost as the size of the system increases. A possible alternative is encoding those pure states into Matrix Product States [162, 163], which compress more efficiently the information and allow for cheap evaluation of quantities such as entanglement entropy [164].

Alternative approaches forego an exact description of the pure state, and perform a semiclassical approximation of the system to obtain a set of $\mathbb{C}-$number SDEs. While this approximation is uncontrolled, the regime is often of interest in quantum optical experiments, and it reduces the computational cost exponentially.

The first approach of this kind was proposed by Vogel and Risken in 1989 [112] and was based upon a truncation of the phase-space master equation, dubbed Truncated Wigner Approximation.

Truncated Wigner Approximation

The derivation of this method depends on the particular system chosen. We will introduce the approximation in the context of a single-mode Kerr resonator with single photon losses

and a coherent drive, described by the Hamiltonian (written in the frame rotating at the pump frequency and assuming $\hbar = 1$),

$$\hat{H} = -\Delta \hat{a}^\dagger \hat{a} + \frac{U}{2} \hat{a}^\dagger \hat{a}^\dagger \hat{a}\hat{a} + F(\hat{a}^\dagger + \hat{a}), \tag{3.27}$$

with jump operator $\mathcal{D} = \sqrt{\gamma}\hat{a}$. The Lindblad master equation

$$\frac{d\hat{\rho}}{dt} = -i\Big[\hat{H}, \hat{\rho}\Big] + \frac{\gamma}{2}\Big\{\hat{\rho}, \hat{a}^\dagger \hat{a}\Big\} - \gamma \hat{a}\hat{\rho}\hat{a}^\dagger, \tag{3.28}$$

can be rewritten as a master equation for the quasi-probability distribution $W_s(\alpha, \alpha^\star)$ using the rules 2.71 derived in Section 2.3. The time-evolution of the Wigner ($s = 0$) Housimi-Q ($s = 1$) and Glauber-p ($s = -1$) representations are encoded into the following partial differential equation

$$\frac{\partial W_s(\alpha, \alpha^\star)}{\partial t} = L_s W(\alpha, \alpha^\star) \tag{3.29}$$

where the linear operator L_s is given by

$$L_s = -\frac{\partial}{\partial \alpha}\left\{\Big[\ \ i\Big(\Delta - U(|\alpha|^2 - 1)\Big) - \gamma\Big]\alpha\ \ + iF\right\} \tag{3.30}$$

$$-\frac{\partial}{\partial \alpha^\star}\left\{\Big[-i\Big(\Delta - U(|\alpha|^2 - 1)\Big) - \gamma\Big]\alpha^\star - iF\right\} \tag{3.31}$$

$$+\ \ \gamma(1-s)\frac{\partial^2}{\partial\alpha\partial\alpha^\star} \tag{3.32}$$

$$+ is\frac{U}{2}\left(\frac{\partial}{\partial\alpha^2}\alpha^2 - \frac{\partial}{\partial\alpha^{\star 2}}\alpha^{\star 2}\right) \tag{3.33}$$

$$+ iU(1-s^2)\frac{\partial^2}{\partial\alpha\partial\alpha^\star}\left(\frac{\partial}{\partial\alpha}\alpha - \frac{\partial}{\partial\alpha^\star}\alpha^\star\right). \tag{3.34}$$

The operator L_s constitutes a generalised Fokker-Planck equation, as derivatives with order higher than two appear. As this operator encodes the whole complexity of the Liouvillian $\mathcal{L}$, solving the time-dynamics of Eq. (3.29) is at least as complex as solving the Lindblad Master Equation (in practice, solving the PDE is an even more complicated problem from the numerical point of view). To lower the computational complexity, we could exploit the fact that a Fokker-Planck equation can be unravelled into a stochastic differential equation for some stochastic trajectories, and compute the expectation values by stochastically sampling only a certain number of trajectories. Unfortunately, there is no established and well-regarded technique to perform such unravelling when terms such as those in Eq. (3.34) are present, as they cannot be mapped onto a drift or diffusion term of a Langevin process uniquely [113]. Instead, one usually writes the most general Langevin noise term and computes the associated Fokker-Planck equation. Third order terms, which arise from the (anti)-normal ordering, could be eliminated by setting $s = \pm 1$ and choosing the P or Q representation. However, we would be left to deal with the term of Eq. (3.33), which reads:

$$\propto is\int\left(\frac{\partial}{\partial\alpha^2}\alpha^2 - \frac{\partial}{\partial\alpha^{\star 2}}\alpha^{\star 2}\right)W_s(\alpha, \alpha^\star)d\alpha d\alpha^\star. \tag{3.35}$$

Such term can in general take non-positive values, thus leading to numerical divergences when trying to integrate the associated SDE trajectories [114, 116].

The Truncated Wigner Approach takes instead a different path: by considering the Wigner representation ($s = 0$) the term of Eq. (3.33) leading to non positive diffusion is excluded (it is simple to verify that Eq. (3.32), being a positive diagonal term, leads to a constant positive diffusion). Lastly, assuming that $U \ll \gamma$, one can truncate the third-order derivates, obtaining the valid, positive definite FP equation for the Wigner probability

$$\frac{\partial W(\alpha,\alpha^\star)}{\partial t} = \frac{\partial}{\partial \alpha} D(\alpha,\alpha^\star) W(\alpha,\alpha^\star) + \frac{\partial}{\partial \alpha^\star} D^\star(\alpha,\alpha^\star) W(\alpha,\alpha^\star) + \Sigma \frac{\partial^2}{\partial \alpha \partial \alpha^\star} W(\alpha,\alpha^\star), \quad (3.36)$$

where γ is the diagonal positive diffusion constant and the drift operator $D(\alpha,\alpha^\star)$ is given by

$$D(\alpha,\alpha^\star) = \left\{ \left[i\left(\Delta - U(|\alpha|^2 - 1)\right) - \gamma \right] \alpha + iF \right\}. \quad (3.37)$$

For FP equations with diagonal constant diffusion, the associated SDE is simply the Langevin equation [113] given by

$$d\alpha(t) = -D(\alpha,\alpha^\star) + \sqrt{\frac{\gamma}{2}} dW(t) \quad (3.38)$$

$$= \left\{ \left[i\left(\Delta - U(|\alpha|^2 - 1)\right) - \gamma \right] \alpha(t) + iF \right\} dt + \sqrt{\frac{\gamma}{2}} dW(t), \quad (3.39)$$

where $dW(t)$ is a Wiener process. This SDE can easily be generalised to systems with many coupled bosonic modes, where there will be a diagonal noise for each mode. The numerical methods to solve this kind of SDE have already been discussed at the end of Section 3.4.2.

We point out that to sample the symmetric functions $O(\alpha,\alpha^\star)$, corrseponding to some operator $\hat{O}$ one can use the simple identity

$$\langle O \rangle = \int O(\alpha,\alpha^\star) W(\alpha,\alpha^\star) d\alpha d\alpha^\star = \frac{1}{N} \sum_{i=1}^{N} O(\alpha_i(t), \alpha_i^\star(t)) \quad (3.40)$$

for the sample of N trajectories $\{\alpha_i(t)\}$. While evaluating observables requires somewhat a small number of trajectories even in systems with many sites, a limit of those unravelling techniques is that it is very hard to compute the Von-Neumann entropy for more than a few sites, as it requires computing the integral of the log-probability across the whole phase space.

Coherent state trajectories

The approach has recently been generalised by Verstraelen and Wouters [124], who showed that by projecting the SDE of photon-counting (Eq. (3.22)) and homodyne (Eq. (3.26)) trajectories onto the basis of gaussian states. To apply the technique a choice of basis, such as the two quadratures X, P or the phase and amplitude N, Φ must be chosen and gaussianity of the two conjugate variables is imposed. The technique has a computational

complexity comparable to that of the Truncated Wigner[5] , retaining the continuous monitoring interpretation of the trajectories, but depending on the system and the chosen set of basis, it can give less accurate results than the Truncated Wigner Approximation.

3.5 Disordered Systems

It is often true that to better understand the physics of a system it is useful to simplify to a great extent the mathematical model used to describe it, such as assuming spatial homogeneity. But once the underlying physical properties and phase diagram of a system are understood, the scenario can be further enriched if one considers the effects of disorder. For closed systems, the interplay of interaction and disorder in a quantum many-body system can lead to remarkable effects such as the breakdown of ergodicity due to quantum effects, a phenomenon known as many-body localisation [165, 166, 167]. Moreover, disorder and glass dynamics are believed to play a fundamental role in the high encoding capability of Spin Glasses (Neural Networks). Some studies on the subject have recently been performed in the context of open systems [168, 169, 170, 171, 172, 173, 174] in order to unveil the competition between dissipation and localization phenomena. However, the theoretical investigations of the nature of far-from-equilibrium correlated phases in the presence of disorder have only scratched the surface [175]. Moreover, while the non-equilibrium extension of stochastic mean-field theory has been developed [176], the scarcity of effective numerical and analytical tools to tackle the problem beyond the mean-field paradigm has limited the study of the effects of disorder on extended systems. Indeed, dealing with disordered systems introduces an additional layer of complexity with respect to the clean problem because it requires to average over a number of disorder realisations to extract the properties of the system. This motivated us to develop an efficient numerical approach to simulate the dynamics of generic open many-body quantum systems in presence of disorder.

In Section 3.4 we have shown that a convenient way to study open quantum systems are stochastic quantum trajectories, which considerably lower the memory requirements compared to density-matrix based methods, while reducing the wall-clock time necessary to carry out the simulations because those calculations can be easily distributed across several computing nodes.

In this section we will first briefly introduce our notation for dealing with disordered systems (Section 3.5.1), and then we will present an optimal stochastic trajectory approach to efficiently compute observables of disordered systems (Section 3.5.3) which we have proposed in a recent publication [2].

3.5.1 Disordered open quantum systems

In a system with static disorder, we take into account random variations of the parameters describing the system, which we encode in the vector of random variables $\mathbf{w}$. Those inhomogeneities satisfy a certain probability distribution $p_D(\mathbf{w})$, which will determine how likely a certain configuration $\mathbf{w}$ is to occur. The open system dynamics of the disordered

[5] As truncated Wigner trajectories tend to fluctuate wildly, as opposed to gaussian trajectories, the authors claim that this technique requires many fewer trajectories to get a sensible averaging.

system is obtained by adding to the Lindbladian of the clean system $\mathcal{L}_{\text{clean}}$ a second Lindblad term $\mathcal{L}_D(\mathbf{w})$, which accounts for both coherent and incoherent disorder-induced processes

$$\mathcal{L}(\mathbf{w}) = \mathcal{L}_{\text{clean}} + \mathcal{L}_D(\mathbf{w}). \tag{3.41}$$

Let us note that $\mathcal{L}(\mathbf{w})$ is still a well-defined Lindbladian according to Eq. (2.27) and can be formally integrated to give the time-evolution of the density matrix for the system in the presence of static disorder,

$$\hat{\rho}(t, \mathbf{w}) = \mathrm{e}^{\mathcal{L}(\mathbf{w})t}\hat{\rho}_0, \tag{3.42}$$

where $\hat{\rho}_0$ is the state at $t = 0$. The expectation value of an observable $\hat{O}$ at a given time t for a specific disorder realisation $\mathbf{w}$ reads as

$$\langle\hat{O}\rangle(t, \mathbf{w}) = \mathrm{Tr}\left[\hat{O}\rho(t, \mathbf{w})\right]. \tag{3.43}$$

In order to access the statistical properties of the system we need to take the *average* of observables over the possible disordered configurations $\mathbf{w}$, weighted by their probability $p_D(\mathbf{w})$. We will refer to this average as the *configuration* average, and will represent it with an overline $\overline{\bullet}$. Dropping the time-dependance we get:

$$\overline{\langle\hat{O}\rangle} = \int d\mathbf{w}\ p_D(\mathbf{w})\ \langle\hat{O}\rangle(\mathbf{w}) \tag{3.44}$$

$$= \lim_{R\to\infty} \frac{1}{R}\sum_{i=1}^{R}\langle\hat{O}\rangle(\mathbf{w}_i), \tag{3.45}$$

where, in the second line, we introduced the sum over the set of R disorder configurations $\{\mathbf{w}_i\}_{i=1,\dots,R}$ sampled according to the distribution $p_D(\mathbf{w})$.

Note that it is always possible to define a disorder-averaged density matrix as

$$\overline{\hat{\rho}}(t) = \left[\int d\mathbf{w}\ p_D(\mathbf{w})e^{\mathcal{L}(\mathbf{w})t}\right]\hat{\rho}_0, \tag{3.46}$$

such that $\mathrm{Tr}\left[\overline{\hat{\rho}}(t)\hat{O}\right] = \overline{\langle\hat{O}\rangle}$ for any given observable $\hat{O}$. Through Eq. (3.46), it is also possible to define the averaged (or quenched) propagator

$$\exp[\mathcal{L}_{\text{ave}}t] = \int d\mathbf{w}\ p_D(\mathbf{w})e^{\mathcal{L}(\mathbf{w})t}, \tag{3.47}$$

which governs the averaged dynamics of the system. This would allow us to obtain directly the disorder-averaged dynamics of the system in the particular cases when the formal expression of Eq. (3.47) can be exactly integrated.

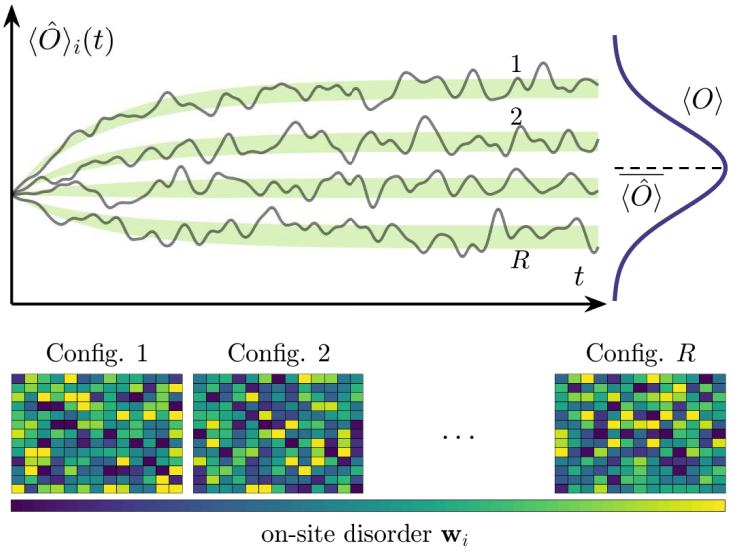

Figure 3.3: Sketch of the optimal sampling method for a disordered open system, whose dynamics is unraveled in terms of individual stochastic trajectories. The configuration-averaged expectation value $\overline{\langle\hat{O}\rangle}(t)$ of an observable is obtained by averaging its value $\left\langle\hat{O}\middle|\hat{O}\right\rangle_i(t)$ over an ensemble of trajectories $i = 1, 2, \dots R$, where each individual trajectory is evolved in the presence of a different disorder configuration $\mathbf{w}_i$.

3.5.2 Density-matrix approach to disordered systems

Obtaining the analytic form of the disorder-averaged propagator is a formidable task, and in general cannot be relied upon. Therefore, to treat disordered systems one has to resort to *manually* performing the disorder average of several density matrices evolved according to Eq. (2.27). One may compute the Liouvillians $\mathcal{L}(\mathbf{w}_i)$ for a set of R disorder configurations $\{\mathbf{w}_i\}_{i=1,\dots,R}$, and then compute the density matrices $\hat{\rho}(\mathbf{w}_i)$ with one of the methods presented in Sections 3.1 to 3.3. Only at the end one can average the density matrices, obtaining $\bar{\hat{\rho}}$. This approach increases the computational cost of studying a disordered system by R-times. Indeed, each density matrix $\hat{\rho}$ has dimension $d^N \times d^N$ for a system made of N sites with local Hilbert space having dimension d, and the method suffers from the same memory limits as methods for clean systems.

3.5.3 Optimal stochastic unraveling of disordered open quantum systems

As mentioned in Section 3.4, numerical methods based on quantum trajectories lower the memory cost from $\mathcal{O}(d^{2N})$ to $\mathcal{O}(d^N)$ [148]. In this section we present our approach, sketched in Fig.3.3, to the problem of disordered open quantum systems.

Statistical errors of stochastic trajectories

We have already shown that the master equation can be formally written as an integral over the space of trajectories

$$\hat{\rho} = \int_{\mathcal{H}} d\psi \; p(\psi) \, |\psi\rangle \, \langle\psi| \, , \tag{3.48}$$

which is the continuous limit of Eq. (3.20) where we omitted the time-dependence. The distribution $p(\psi)$ is sampled by independently evolving a finite number of trajectories. As this is a stochastic sampling of the distribution, the error of the estimated expectation value $\hat{O}$ scales as [6]

$$\epsilon_{\text{clean}} = \alpha \sqrt{\frac{\mathbb{V}\text{ar}_\psi[\hat{O}]}{T}}, \tag{3.49}$$

where T is the number of trajectories and $\mathbb{V}\text{ar}_\psi[\hat{O}]$ is the statistical variance of the observable across the trajectories which depends on the specific evolution protocol and on the observable $\hat{O}$. The prefactor $\alpha = \epsilon/\sigma$ arises from the definition of the absolute deviation $\epsilon = \langle|\hat{O} - \langle\hat{O}\rangle|\rangle$, as opposed to that of the standard deviation $\sigma = \sqrt{\langle\hat{O}^2 - \langle\hat{O}\rangle^2\rangle}$. As it is obvious from the definition, both are measures of the deviation (spread) of the distribution $\mathcal{O}$, but while the first has a more *direct* interpretation when sampling, the latter has much better mathematical properties. It is possible to show that $\alpha \in [0, 1]$, and we will simply treat it as a trivial proportionality factor with a value usually around that of the normal distribution, namely $\alpha = \sqrt{2/\pi}$ [177, pgs. 81-85 and 113].

Stochastic trajectories for disordered systems

We can now introduce our method to perform an efficient estimate of disorder-averaged expectation values. Our aim is to efficiently compute the expectation value defined by Eq. (3.44) when the density matrix is unravelled stochastically. By inserting Eq. (3.48) into Eq. (3.44) we get:

$$\overline{\langle\hat{O}\rangle} = \int d\mathbf{w} \; p_D(\mathbf{w}) \int d\psi \; p(\psi|\mathbf{w}) o(\psi), \tag{3.50}$$

where $p(\psi|\mathbf{w}) = p(\psi, \mathbf{w})/p_D(\mathbf{w})$ is the conditional probability of obtaining the state $|\psi\rangle$ given the disorder configuration $\mathbf{w}$ and $o(\psi) = \langle\psi(\mathbf{w})| \, \hat{O} \, |\psi(\mathbf{w})\rangle$. Sampling the disorder-averaged expectation value can therefore be interpreted as performing the sampling of multivariate distribution $p(\psi, \mathbf{w})$ which depends on ψ and $\mathbf{w}$.

There exist several ways to perform such a sampling. Indeed, we have freedom to choose how many trajectories we want to evolve according to a given disorder configuration. In the general case one evolves T trajectories for each of the R disorder realisations for a total computational cost $\mathcal{C} = R \times T$. We will show that, when computing disorder-averaged

[6] This assumes that all trajectories are independent. Because the initial state is fixed, for small times below the correlation time the trajectories are not independent variables, and a covariance correction to the variance should be included through the substitution $\mathbb{V}\text{ar}_\psi \to \mathbb{V}\text{ar}_\psi + \mathbb{C}\text{ov}_\psi$. As this is related to the correlation time, it must decay exponentially in time. Note that this only affects the efficiency of the sampling, but not the expectation value, which is unaffected.

expectation values, the most computationally efficient sampling strategy is realised for $T = 1$, i.e. one needs to evolve only one trajectory for each disorder configuration. This strategy is sketched in Fig. 3.3. In this case, the error committed when computing the expectation value is

$$\epsilon^2 = \alpha^2 \frac{\overline{\langle \hat{O}^2 \rangle} - \left(\overline{\langle \hat{O} \rangle} \right)^2}{R} \tag{3.51}$$

$$= \alpha^2 \frac{\overline{\langle \hat{O}^2 \rangle - \langle \hat{O} \rangle^2} + \left[\overline{\langle \hat{O} \rangle^2} - \left(\overline{\langle \hat{O} \rangle} \right)^2 \right]}{R}. \tag{3.52}$$

The first term $\overline{\langle \hat{O}^2 \rangle - \langle \hat{O} \rangle^2} = \overline{\mathbb{V}\mathrm{ar}_\psi[\hat{O}]}$ is the disorder-averaged statistical variance of the observable [see Eq. (3.49)] and accounts for the uncertainty due to the trajectory protocol. This is the same contribution found in the clean case described by Eq. (3.49). The second term $\overline{\langle \hat{O} \rangle^2} - \left(\overline{\langle \hat{O} \rangle} \right)^2 = \mathbb{V}\mathrm{ar}_D[\langle \hat{O} \rangle]$, instead, is the variance of the expectation value of the observable $\langle \hat{O} \rangle$ due to the presence of disorder. Exploiting these identifications, we can rewrite Eq. (3.52) in a more compact form as

$$\epsilon = \alpha \sqrt{\frac{\overline{\mathbb{V}\mathrm{ar}_\psi(\hat{O})} + \mathbb{V}\mathrm{ar}_D(\langle \hat{O} \rangle)}{R}}, \tag{3.53}$$

which is valid when sampling one trajectory for each disorder realisation. The generalisation of Eq. (3.53) to the $T > 1$ case is obtained by performing the substitution $\mathbb{V}\mathrm{ar}_\psi[\hat{O}] \to \mathbb{V}\mathrm{ar}_\psi[\hat{O}]/T$ which reflects the reduction in the statistical error when evolving multiple trajectories. After some algebra, we obtain the general formula for the error as a function of R and T

$$\epsilon(R, T) = \alpha \sqrt{\frac{\overline{\mathbb{V}\mathrm{ar}_\psi(\hat{O})} + T\, \mathbb{V}\mathrm{ar}_D(\langle \hat{O} \rangle)}{RT}}. \tag{3.54}$$

We point out that in the absence of disorder, $\mathbb{V}\mathrm{ar}_D = 0$ and the only term playing a role is the uncertainty associated with the trajectories. In this case, the distinction between different disorder configurations becomes meaningless, and Eq. (3.54) reduces to Eq. (3.49) where the total number of trajectories is $T \to \mathcal{C}$. Equivalently, when integrating the master equation exactly for R different configurations of $\mathbf{w}$ one obtains a similar formula where the contribution of the trajectories is absent ($\mathbb{V}\mathrm{ar}_\psi = 0$) and only $\mathbb{V}\mathrm{ar}_D$ contributes to the error.

From Eq. (3.54) it is easy to deduce that the most efficient sampling strategy, i.e. the choice of T which minimises the error for a fixed computational cost $\mathcal{C} = R \times T$, is realised for $T = 1$. We remark that the only assumption we made to obtain Eq. (3.54) is to consider that the dynamics of the T trajectories we evolve for each disorder realisation is not correlated [7], as required for an effective stochastic unraveling [85]. Thus Eq. (3.54) holds for arbitrary evolution protocols and disorder distributions.

[7] If the evolution is started from a pure state, then the trajectories will be correlated for times $t < \tau$ below the correlation time. Nevertheless, our method still yields correct results, but the error will scale slightly worse, as there are cross-correlations that are harder to average out.

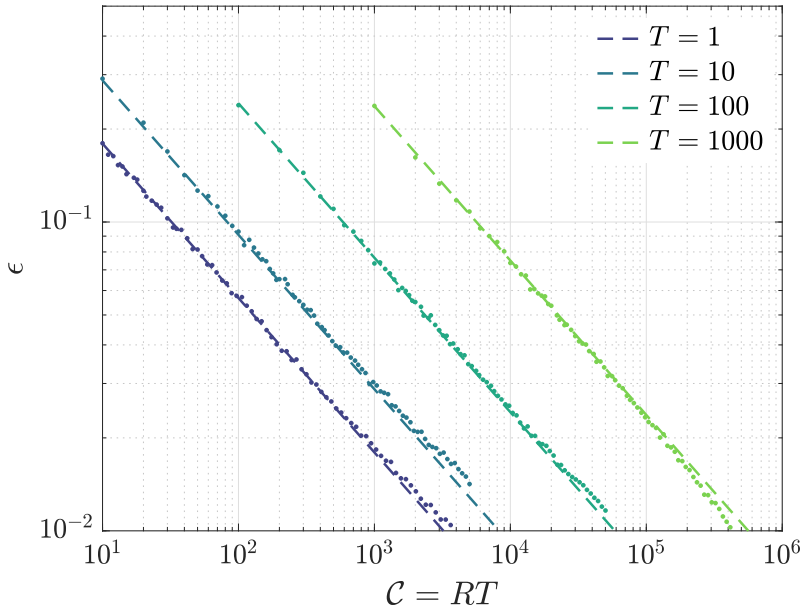

Figure 3.4: Mean absolute error $\epsilon = |n_{ss} - n_{\text{exact}}|/2$ as a function of the computational cost $\mathcal{C}$, for the average number of bosons in a driven-dissipative Bose-Hubbard chain made of 5 sites with $U = 1.0\gamma$, $F = 2.0\gamma$, $J = 0.5\gamma$ and $\Delta = \mathcal{N}(1.0, 0.5)$. The dashed lines are predicted by $\epsilon(\mathcal{C}/T, T)$ as defined in Eq. (3.54), while dots are numerical data. The cases of $T = 1, 10, 100, 1000$ trajectories are considered. Numerical data has been computed using a combination of *QuTiP* [152, 153] and *QuantumOptics.jl* [154].

3.5.4 Numerical verification

As a benchmark of the optimality of the method, we have applied the optimal stochastic unraveling to the driven-dissipative Bose-Hubbard model with on-site static disorder. Here we simply use system as a benchmark. If the reader is interested in the physical details of this system, he can refer to Chapter 4. The coherent part of the evolution (in the frame rotating at the driving frequency and setting $\hbar = 1$) is ruled by the following Hamiltonian

$$\hat{H}_{\text{BH}} = \sum_{j=1}^{L} \left[-\Delta_j \hat{a}_j^\dagger \hat{a}_j + \frac{U}{2} \hat{a}_j^{\dagger 2} \hat{a}_j^2 + F \left(\hat{a}_j^\dagger + \hat{a}_j \right) \right] - J \sum_{\langle i,j \rangle} \hat{a}_i^\dagger \hat{a}_j, \tag{3.55}$$

where the local energies $\{\Delta_j\}$ are assumed to be random variables obeying a gaussian distribution $\mathcal{N}(\Delta_0, W)$ with mean Δ_0 and variance W^2. Dissipative processes, i.e. local boson leakage, are modelled with a set of local jump operators $\hat{L}_j = \hat{a}_j$ (j being the site index) with uniform rate γ leading to the following expression for the master equation

$$\partial_t \hat{\rho}(t) = -i \left[\hat{H}_{\text{BH}}, \hat{\rho} \right] + \frac{\gamma}{2} \sum_{j=1}^{L} \left[2\hat{a}_j \hat{\rho} \hat{a}_j^\dagger - \left\{ \hat{a}_j^\dagger \hat{a}_j, \hat{\rho} \right\} \right]. \tag{3.56}$$

We benchmark against a chain of $L = 5$ sites with periodic boundary conditions and

interaction strength $U = \gamma$. First, we have computed the observables via a brute force integration of the master equation averaged over 2000 disorder configurations, sufficient to reduce the statistical error below $5 \cdot 10^{-3}$. We take the value of the average boson density $\hat{n} = \sum_{j=1}^{L} \hat{a}_j^\dagger \hat{a}_j / L$ computed in this way as the *exact* value n_{exact} against which we will compare the sampling performed with trajectories.

We then unravel the dynamics of the master equation (3.56) in terms of photon-counting quantum trajectories evolved according to the wavefunction Monte Carlo method presented in Section 3.4.1 [96, 151]. We generate R disorder configurations $\mathbf{w}$, and compute $T = 1,\ 10,\ 100,\ 1000$ different trajectories for every configuration. We extract the steady-state average boson density n_i for every $i = 1, \ldots, \mathcal{C} = RT$ trajectory.

In Fig. 3.4 we show the mean absolute error $\epsilon(R,T) = \frac{1}{\mathcal{C}} \sum_{i=1}^{\mathcal{C}} |n_i - n_{exact}|$ computed both numerically and analytically (Eq. (3.54) with $\alpha = \sqrt{2/\pi}$), finding a perfect agreement. The fact that at a fixed computational cost $\mathcal{C}$ the lowest sampling error ϵ is achieved by considering $T = 1$ further confirms the optimality of our method.

3.6 Conclusions

To summarise, we have shown that there exist several numerical methods to treat finite size open quantum systems, which can often be combined together. In general, if one is only interested in the steady-state, this can be computed with a much lower computational cost compared to the full time-evolution. Iterative solvers targeting the steady-state can be combined with other ansätze that truncate that space, either semiclassically such as Gützwiller or by computing the corner space. Alternatively, an efficient and scalable approach is ensured by quantum trajectories, which can also be combined with several ansätze. The main limitation for those numerical methods is the entropy of the state, which limits the effectiveness of rank-reduction techniques, and the presence of quantum correlations, which cannot be treated semiclassically. Tensor network techniques, which perform remarkably for very mixed states, are limited to only 1 dimensional systems, and only few proof of concept exist in higher dimensional systems.

At the end of the chapter we have also introduced a general sampling method to efficiently compute configuration averages of observables in disordered open systems, whose dynamics has been unraveled with stochastic trajectories. We have proven that the optimal strategy is realised when each trajectory evolves according to a different disorder configuration. Our approach allows us to drastically reduce the computational cost needed to perform disorder-averaged quantities, because the scaling of the statistical error with the number of trajectories and disorder configurations is independent of the specific stochastic evolution protocol and of the particular disorder distribution

CHAPTER 4

Critical phenomena in the driven-dissipative Bose-Hubbard model

In a closed many-body quantum system at zero temperature, the pure ground state may undergo a quantum phase transition when there is a competition between two physical processes described by non-commuting Hamiltonian terms [178]. In an open system, the competition between unitary Hamiltonian evolution and incoherent dissipation can induce a dissipative phase transition for the steady-state in the thermodynamic limit [39], as it has been discussed theoretically for photonic systems [179, 180, 181, 182, 183], lossy polariton condensates [184, 185, 186] and spin models [43, 187].

In this chapter we will present a thorough investigation of the the driven-dissipative Bose-Hubbard model, and the closely-related phenomenon of optical bistability. We will begin in Section 4.1 by briefly reviewing the experimental platforms relevant for this model. In Section 4.2 we will introduce the physics of a single dissipative nonlinear resonator. In Section 4.2.4 we will show that by coherently driving several coupled resonators it is possible to create a macroscopic population of bosons in the zero-wavevector mode ($k = 0$). In Section 4.3 we will show via numerical results that if those resonators exhibit non-linear behaviour, this mode can be linked with a first-order phase transition that remarkably exhibits long-range order. Moreover, we will show that such first-order criticalities are signalled by the emergence of a bistable behaviour. In Section 4.4 we will investigate the effects of disorder. This theoretical investigation is relevant to experiments on lattices of photonic resonators can be constructed with relative ease [26, 188, 189, 190], in platforms such as patterned semiconductor microcavities [191] and nonlinear superconducting microwave resonators [24, 192, 193].

Contents

4.1 Experimental systems

In this section we will thoroughly study the physics of interacting massive bosonic particles in a dissipative context. There exist several experimental platforms that allow the confinement and control of bosonic particles, such as optomechanical resonators [27], Rydyberg atoms [187, 194, 195], exciton polaritons [30] and superconducting circuits [193].

4.1.1 Exciton polaritons

Particularly relevant experimental platforms are those exploiting hybrid light-matter quasiparticles, the so-called polaritons. The underlying idea is that confining light in a cavity gives it an effective mass, and photon-photon interaction can be mediated by a matter degree of freedom strongly coupled to the photons. The confinement of photons is achieved by delimiting a cavity through some reflective mirrors. At optical frequencies stacks of Bragg Mirrors are used, which allow to create quasi 0-dimensional, 1D or 2D electromagnetic cavities. 2D cavities are interesting because the in-plane (parallel to the mirrors) dispersion relation of the confined light reproduces that of a massive particle. This can be seen by expanding to second order in $|k_\perp|/k_z$ the modulus of the wave-vector $\mathbf{k} = \{\mathbf{k}_\perp,\ k_z\}$,

$$|\mathbf{k}| = \sqrt{k_z^2 + |\mathbf{k}_\perp|^2} = k_z \left(1 + \frac{|\mathbf{k}_\perp|^2}{2k_z^2}\right) + \mathcal{O}(\frac{|k_\perp|^4}{k_z^4}). \tag{4.1}$$

and by substituting this term into the dispersion relation

$$\omega_{\rm ph}(\mathbf{k}) = \frac{\hbar c}{n_0}|\mathbf{k}| = m_{\rm eff}c^2 + \frac{(\hbar|\mathbf{k}_\perp|)^2}{2m_{\rm eff}}, \tag{4.2}$$

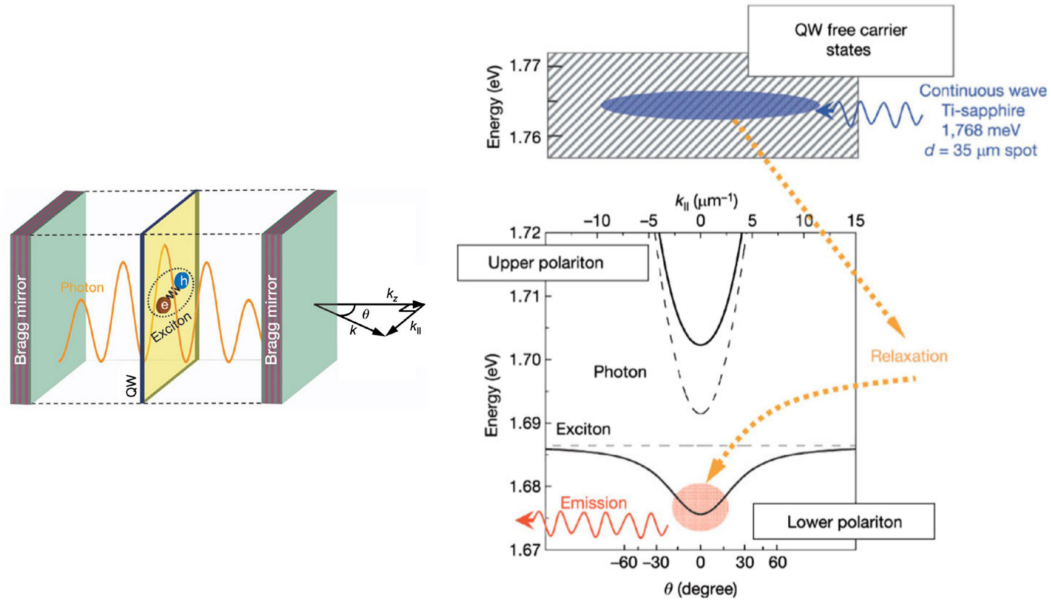

Figure 4.1: Left panel: Sketch of a planar semiconductor microcavity delimited by two Bragg mirrors and embedding a quantum well (QW). The wave vector in the z direction perpendicular to the cavity plane is quantized, while the in-plane motion is free. The cavity photon mode is strongly coupled to the excitonic transition in the QW. Right panel: energy dispersion for an exciton-polariton. In the lower panel the 2D photonic dispersion and the excitonic part are represented with dashed lines. The lower and upper polariton dispersions are drawn with continuous lines. In the top panel a continuum excitonic modes are depicted, which when pumped will relax towards the lower polariton. From [26].

where n_0 is the refraction index of the medium between the mirrors, k_z is quantised due to the boundary conditions imposed by the mirrors, and $m_{\text{eff}} = \hbar n_0 k_z / c$ is the effective mass of the photons confined to the 2D in-plane space.

The non-linear interaction among photons, which is usually negligible at optical frequencies in the vacuum [34], can be induced by inserting a nonlinear medium, such as a quantum well, inside of the Fabry-Perot cavity. When light is shined on semiconductors, photons might be absorbed by electrons, which are then promoted from the valence band into the conduction band. When this happens an electron-hole pair is created; as those particles have opposite charge, they attract each other, binding together into a quasi-particle with bosonic statistics, the exciton. Excitons interact through the Coulomb force because of their fundamental electronic nature. and they decay radiatively by emitting a photon in the cavity.

This system can be modelled theoretically by the Hamiltonian for the cavity photons ($\hat{c}_{\mathbf{k}}$) and excitons ($\hat{d}_{\mathbf{k}}$),

$$\hat{H} = \sum_{\mathbf{k}} \hbar \left[\omega_{\text{ph}}(\mathbf{k}) \hat{c}^{\dagger}_{\mathbf{k}} \hat{c}_{\mathbf{k}} + \omega_{\text{ex}}(\mathbf{k}) \hat{d}^{\dagger}_{\mathbf{k}} \hat{d}_{\mathbf{k}} + \Omega(\hat{d}^{\dagger}_{\mathbf{k}} \hat{c}_{\mathbf{k}} + \hat{c}^{\dagger}_{\mathbf{k}} \hat{d}_{\mathbf{k}}) \right] + \sum_{\mathbf{k},\mathbf{k}',\mathbf{q}} \frac{U_{\text{ex-ex}}}{2} \hat{d}^{\dagger}_{\mathbf{k}+\mathbf{q}} \hat{d}^{\dagger}_{\mathbf{k}'-\mathbf{q}} \hat{d}_{\mathbf{k}} \hat{d}_{\mathbf{k}'}, \tag{4.3}$$

where Ω is the Rabi frequency quantifying the coupling between the bosons and the pho-

tons, and $U_{\text{ex-ex}}$ is the exciton-exciton interaction strength, for which we have omitted the dependency from the wave-vectors. This Hamiltonian can be diagonalised by introducing the polariton operators as a linear combination of the photonic $\hat{c}$ and matter $\hat{d}$ operators [196].

$$\hat{a}_{\mathbf{k}}^{\text{LP/UP}} = u_{\mathbf{k},\ \text{ph}}^{\text{LP/UP}} \hat{c}_{\mathbf{k}} + u_{\mathbf{k},\ \text{ex}}^{\text{LP/UP}} \hat{d}_{\mathbf{k}}. \tag{4.4}$$

The operator $\hat{a}^{\dagger,\text{LP}}$ ($\hat{a}^{\dagger,\text{UP}}$) creates a lower (upper) polariton. For every wave-vector $\mathbf{k}$ the condition $|u_{\mathbf{k},\ \text{ph}}^{\text{LP/UP}}|^2 + |u_{\mathbf{k},\ \text{ex}}^{\text{LP/UP}}|^2 = 1$ implies that $C_{\mathbf{k}} = |u_{\mathbf{k},\ \text{ph}}^{\text{LP/UP}}|^2$ can be interpreted as the photonic fraction of the polariton and $X_{\mathbf{k}} = |u_{\mathbf{k},\ \text{ex}}^{\text{LP/UP}}|^2$ as the excitonic fraction.

With this change of coordinates one can rewrite the Hamiltonian into the form

$$\hat{H} = \sum_{j\in\{\text{LP,UP}\}} \sum_{\mathbf{k}} \hbar\omega_j(\mathbf{k}) \hat{a}_{j,\mathbf{k}}^{\dagger} \hat{a}_{j,\mathbf{k}} + \sum_{\mathbf{k}',\mathbf{q}} \frac{U_{j,j',j'',j'''}}{2} \hat{a}_{j,\mathbf{k}+\mathbf{q}}^{\dagger} \hat{a}_{j',\mathbf{k}'-\mathbf{q}}^{\dagger} \hat{a}_{j'',\mathbf{k}} \hat{a}_{j''',\mathbf{k}'}, \tag{4.5}$$

where $\omega_j(\mathbf{k})$ and $U_{j,j',j'',j'''}$ depend on the coefficients ω_{ph}, ω_{ex}, Ω and $U_{\text{ex-ex}}$. The resulting dispersion relation for a 2D cavity is depicted in Fig. 4.1. If one considers only the lower-polariton branch and considers only a single spatial mode, then the Hamiltonian reads

$$\hat{H}_{\text{single-mode}} = \hbar\omega_0 \hat{a}^{\dagger}\hat{a} + \frac{U}{2} \hat{a}^{\dagger}\hat{a}^{\dagger}\hat{a}\hat{a}, \tag{4.6}$$

where $\omega_0 = m_{\text{eff}}c^2$. this is a Kerr Hamiltonian. The leaking of photons out of the cavity can be modelled within the master-equation formalism by adding single-polariton losses through the dissipative channel $\hat{a}$ with rate γ.

We remark that the polariton-polariton interaction strength for exciton polaritons is usually in the range of $U/\gamma \simeq 10^{-2}$, with a few state of the art devices reported to have nonlinearities as high as $U/\gamma \simeq 5 \cdot 10^{-2}$ [197].

4.1.2 Lattices of resonators

By etching a planar cavity in order to obtain a quasi 0-D structure like a micropillar it is possible to create a system with few and energetically separated modes. The polariton picture can still be used to describe those systems, but instead of a collection of $\mathbf{k}$ modes only few are available. Moreover, by coupling several of those pillars it is possible to engineer 1D chains, 2D lattices or even artificial systems such as benzene molecules [198], which can produce remarkable properties. The coupling is usually achieved by slightly overlapping the pillars, so that the photons, and therefore the polaritons, can hop from one resonator to another (see for example Fig. 4.2). While the exact form of the coupling depends on the shape of the pillars and microscopic details of the system, it is in general possible to model it with a nearest neighbour hopping term of the polariton,

$$H_{hop} = -J \sum_{\langle i,j\rangle} \hat{a}_i^{\dagger}\hat{a}_j + \hat{a}_i\hat{a}_j^{\dagger}, \tag{4.7}$$

where the pedix i labels the site where the creation/destruction operators act, and J is the hopping rate. Devices like that allow to study the physics of interacting lattices, such as the Bose-Hubbard model in the weakly interacting limit.

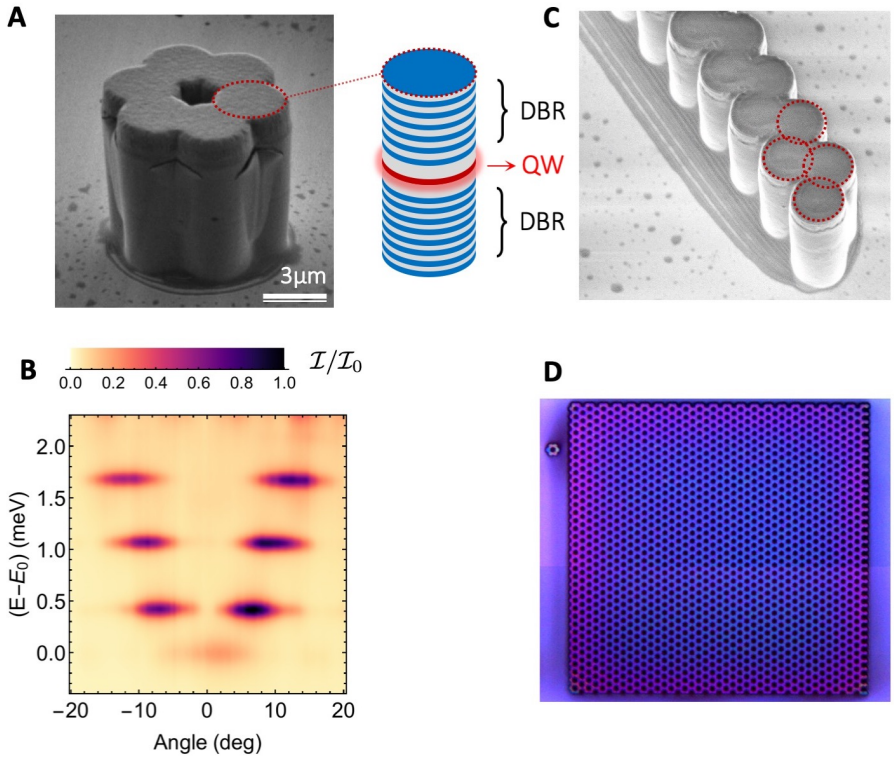

Figure 4.2: **A:** photography of an artificial benzene molecule consisting of 6 coupled micropillars. The structure of the micropillars is sketched on the right, where the top and bottom stacks of Bragg mirrors and the nonlinear quantum well are evidenced. **B:** The experimentally-measured energy dispersion of the structure above. **C-d:** A one-dimensional chain of pillars used to simulate the SSH model, and a two-dimensional honeycomb lattice. Left panels are adapted from [198], while right panels have been kindly provided by the authors.

4.1.3 Superconducting resonators

It is also possible to work with polaritons in the microwave regime using a different physical implementation. At those frequencies the electromagnetic cavity can be implemented via a superconducting transmission line resonator, delimited by two capacitors which behave effectively as the mirrors. Three-dimensional cavities are made by using a superconducting box. The nonlinearity is provided by embedding a nonlinear element, usually a Josephson junction or a SQUID into the resonator. The physics of several coupled resonators can be studied by capacitively (or inductively) coupling the resonators. When working with microwaves it is possible to fabricate resonators with quality factors much higher than the optical counterpart ($10^6 - 10^{10}$ versus $10^3 - 10^5$ for optical setups). Additionally, it is easy to produce devices with very high nonlinearities ($U/\gamma > 10$) [192, 193].

Microwave resonators can be modeled theoretically with the tools of circuits quantum-electro-dynamics, which provide a quantized description of the currents and voltages. In

electrical devices energy can be stored in the electric field of a capacitor, or in the magnetic flux of an inductance. In a capacitance C, the stored energy is given by $E_C = Q^2/2C$ where Q is the accumulated charge. Likewise, the energy stored within an inductance L is given by $E_L = \Phi^2/2L$, where Φ is the magnetic flux. The flux and the charge can be modelled as canonically-conjugated quantum operators with commuting relation $\left[\hat{\Phi}, \hat{Q}\right] = i$. This quantization produces the Hamiltonian

$$\hat{H} = \frac{\hat{\Phi}^2}{2L} + \frac{\hat{Q}^2}{2C}. \tag{4.8}$$

In the same spirit as the quantum harmonic oscillator, it is possible to diagonalise the Hamiltonian by defining the ladder operator, (where $\hbar = 1$),

$$\hat{a} = \frac{1}{\sqrt{\omega}} \left(\frac{\hat{\Phi}}{\sqrt{2L}} - i \frac{\hat{Q}}{\sqrt{2C}} \right), \tag{4.9}$$

where the frequency is given by $\omega = 1/\sqrt{LC}$. With this operator, which respects the traditional commuting relation $\left[\hat{a}, \hat{a}^\dagger\right] = 1$, the Hamiltonian reduces to the standard single-mode harmonic oscillator

$$\hat{H} = \omega(\hat{a}^\dagger \hat{a} + \frac{1}{2}). \tag{4.10}$$

The dissipation is naturally provided by the radiative coupling with the environment, both because of defects in the fabrication of the microwave cavity and because the mirrors are not perfectly reflective.

Interestingly, it is also possible to couple the nonlinear cavity with another degree of freedom, such as an artificial two-level system represented by a Josephson junction. This setup has been used to probe dissipatively stabilized Mott insulators [199], and could be also used to study incoherent pumping schemes [91, 44, 45].

4.2 Single-mode Kerr model

As a pedagogical step towards an understanding of the phenomena that emerges in a lattice of coupled nonlinear cavities, we first briefly discuss the single cavity driven-dissipative Kerr Model.

The Kerr model describes a single bosonic mode with frequency ω_c with boson-boson interaction U, driven by a pump at frequency ω_p and amplitude F. The Hamiltonian of the system is,

$$\hat{H}_{\text{Kerr}} = \omega_c \hat{a}^\dagger \hat{a} + \frac{U}{2} \hat{a}^\dagger \hat{a}^\dagger \hat{a} \hat{a} + iF(e^{i\omega_p t} \hat{a}^\dagger + e^{-i\omega_p t} \hat{a}). \tag{4.11}$$

The coupling to the environment is modelled by single-boson losses at rate γ and jump operator $\hat{L} = \hat{a}$. We shift the frame to the one rotating at the pump frequency with the unitary transformation $U = e^{-i\omega_p t \hat{a}^\dagger \hat{a}}$ in order to obtain the time-independent Hamiltonian

$$\hat{H} = -\Delta \hat{a}^\dagger \hat{a} + \frac{U}{2} \hat{a}^\dagger \hat{a}^\dagger \hat{a} \hat{a} + F(\hat{a}^\dagger + \hat{a}), \tag{4.12}$$

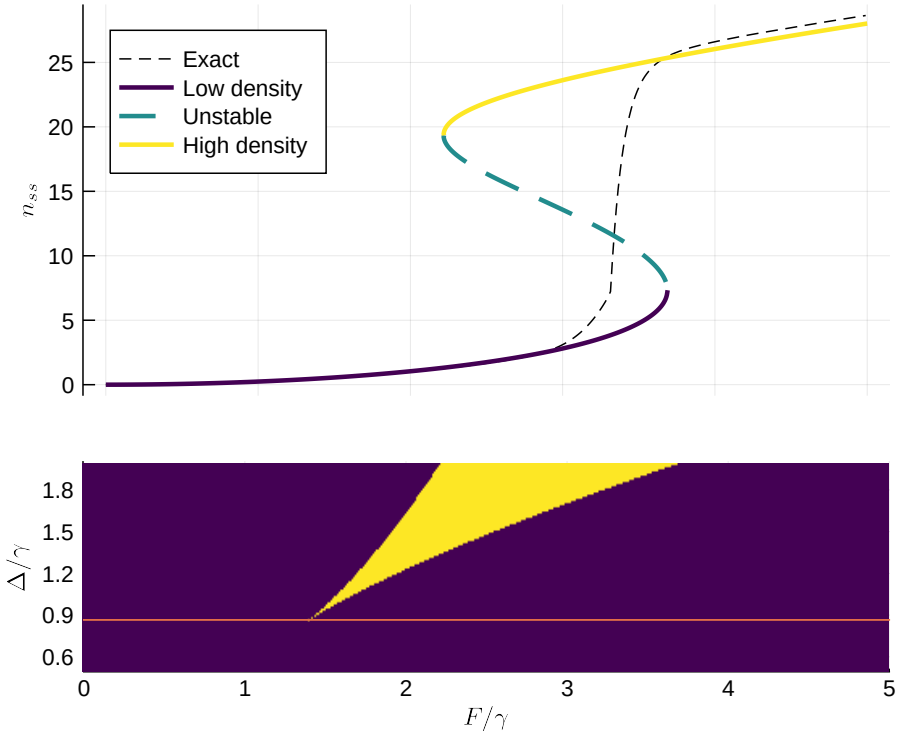

Figure 4.3: Top panel: semiclassical solutions for the steady state average population as a function of the drive amplitude for the Kerr model with detuning $\Delta = 2\gamma$. The bottom line represent the low density solution, the top bright line represent the high density solution and the middle dashed line is the intermediate unstable solution. Bottom panel: number of semiclassical non-degenerate solutions as a function of the drive amplitude F/γ and detuning Δ/γ. The dark area has a single (triple-degenerate) solution while the bright area has three non degenerate solutions. The horizontal line is $\Delta/\gamma = \sqrt{3}/2$. The kerr nonlinearity is $U/\gamma = 0.1$.

where $\Delta = \omega_p - \omega_c$ is the detuning between the cavity and the pump frequency. The resulting Lindblad master equation is therefore

$$\frac{d\hat{\rho}(t)}{dt} = -i\Big[\hat{H}, \hat{\rho}(t)\Big] + \gamma\left(\hat{a}\hat{\rho}(t)\hat{a}^\dagger - \frac{1}{2}\{\hat{a}^\dagger\hat{a}, \hat{\rho}(t)\}\right). \tag{4.13}$$

4.2.1 Semiclassical theory

By computing the equations of motion for the expectation value of $\hat{a}(t)$ operator using the adjoint master equation [21]

$$\frac{d\langle\hat{a}\rangle}{dt} = -i\Delta\langle\hat{a}\rangle + iU\langle\hat{a}^\dagger\hat{a}\hat{a}\rangle + F - \frac{\gamma}{2}\langle\hat{a}\rangle, \tag{4.14}$$

we can then perform a semiclassical analysis by replacing operator-valued expectation values with $\mathbb{C}-$numbers encoding the (classical) coherent state of the system. Upon per-

forming the substitution $\hat{a}(t) \to \alpha(t)$ we obtain the so-called driven-dissipative Gross-Pitaievskii equation

$$\dot{\alpha} = \left\{-i\left[\Delta - U|\alpha|^2\right] - \frac{\gamma}{2}\right\}\alpha + F. \tag{4.15}$$

The steady state equation $\dot{\alpha} = 0$ can be recast into a cubic equation for the mean-field coherent population $n = |\alpha|^2$,

$$n\left[(\Delta - Un)^2 + \frac{\gamma^2}{2}\right] = F^2, \tag{4.16}$$

which has three solutions. It is possible to analytically show that the three solutions are degenerate if $\Delta/\gamma < \Delta^\star/\gamma = \sqrt{3}/2$, while for $\Delta > \Delta^\star$ there always exists an interval $F/\gamma \in [F_<, F_>]$ where the three solutions are non-degenerate, as shown in the bottom panel of Fig. 4.3. By performing a linear stability analysis, one verifies that of the three non-trivial solutions one is actually dynamically unstable. This means that in presence of perturbations or fluctuations the system runs away from the unstable steady state. An example of the three solutions is given in the top panel of Fig. 4.3, where $U/\gamma = 0.1$ and $\Delta/\gamma = 2$.

The existence of two stable solutions at the semiclassical level is in stark contrast to what is given by the full quantum solution found by Drummond and Walls [52], who showed that the system has an unique steady state. The expectation value for the number of photons in the steady state is also plotted in Fig. 4.3. In that figure we also show that the semiclassical theory gives a perfect prediction of the number of photons in the region with a single solution, while it fails in the bistable region.

Unravelling the master equation (4.13) into photon-counted trajectories, as introduced in Section 2.2.2, and computing the expected number of photons $\langle\psi|\,\hat{n}\,|\psi\rangle$ along the resulting stochastic process yields the following: when $F < F_<$ or $F > F_>$ the observed trajectory oscillates around the population predicted by the semiclassical theory. Instead, if F lies in the bistable region, the system oscillates either around the low- or the high-density solutions, and rarely jumps suddenly from one solution to the other.

The two dynamically stable solutions are correctly interpreted as stable by the semiclassical theory. The single solution predicted by the master equation reflects the fact that if the results of the measurements are not kept track of, the system will be in a mixed superposition of the two, and it's impossible to tell in which one it is exactly without measuring.

4.2.2 The driven-dissipative Bose-Hubbard model

Remarkable properties arise in quantum systems when several modes interact. The Bose-Hubbard model, which accounts for particle exchange among several interacting bosonic modes, has been extensively studied in the context of closed Hamiltonian systems both theoretically [200, 201, 202, 203, 204, 205] and experimentally in experiments where ultracold-atoms are loaded in optical lattices [206, 207]. Recent advancements in the fabrication and control of arrays of resonators, both in the microwave and optical frequency domain provides the opportunity to study Bose-Hubbard systems in the presence of driving and dissipation under controlled conditions. After the initial proposal for the experimental

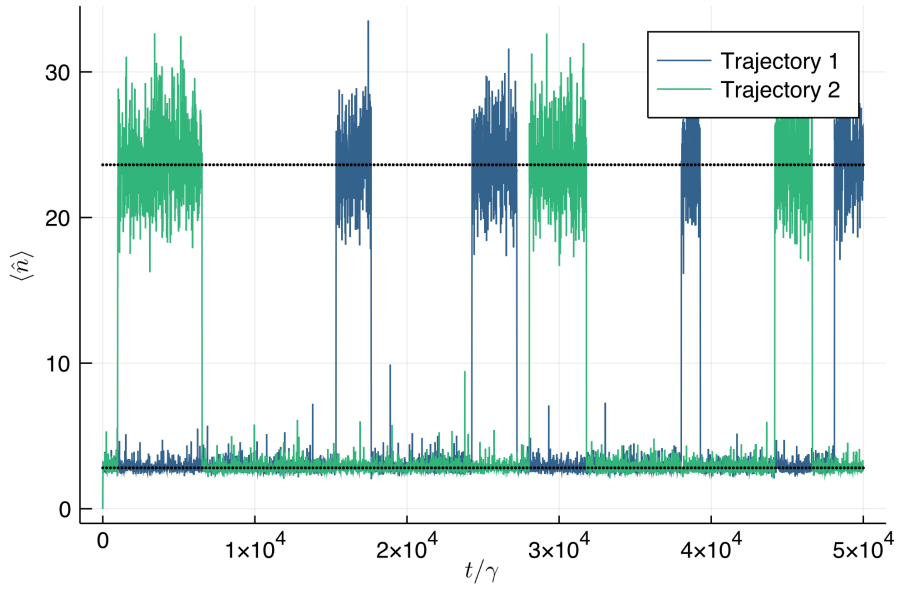

Figure 4.4: Time evolution of the number of photons along two photon-counting trajectories. The semiclassical prediction for the low and high density solutions are represented by dotted lines. Parameters are the same as in the top panel of Fig. 4.3.

implementation in lattices of coupled photonic cavities [208, 209, 210], research initially focused on phenomena close to the Mott insulator-superfluid quantum phase transition [211, 212, 213]. More recently, an extensive investigation of the features found only in the driven-dissipative context, such as bistability and dissipative phase transitions have been carried out both theoretically [26, 214, 215, 181, 183, 216] and experimentally [24, 192, 197].

In the following, we will focus on the driven-dissipative Bose-Hubbard model described by the Hamiltonian,

$$H = -\sum_i \Delta \hat{a}_i^\dagger \hat{a}_i - J \sum_{\langle i,j \rangle} \hat{a}_i^\dagger \hat{a}_j + \hat{a}_i \hat{a}_j^\dagger + \frac{U}{2} \sum_i \hat{a}_i^\dagger \hat{a}_i^\dagger \hat{a}_i \hat{a}_i + \sum_i F(\hat{a}_i^\dagger + \hat{a}_i) \tag{4.17}$$

and with homogeneous single particle loss rate γ on every site, described by the jump operator $\hat{a}_i$.

4.2.3 Mean-field studies

Initial investigations into the lattice model resorted to a combination of mean-field techniques with other approximations. Performing a mean-field approximation on top of the semiclassical approximation described in the previous chapter (setting all fields $\alpha = \alpha_i \; \forall i$) yields the same equations of the Kerr model, where the single-particle energy is replaced by the effective detuning $\Delta \to \Delta + zJ$ (z beeing the coordination number of the lattice), and bistability is still predicted whenever $\Delta + zJ > \sqrt{3}/2$.

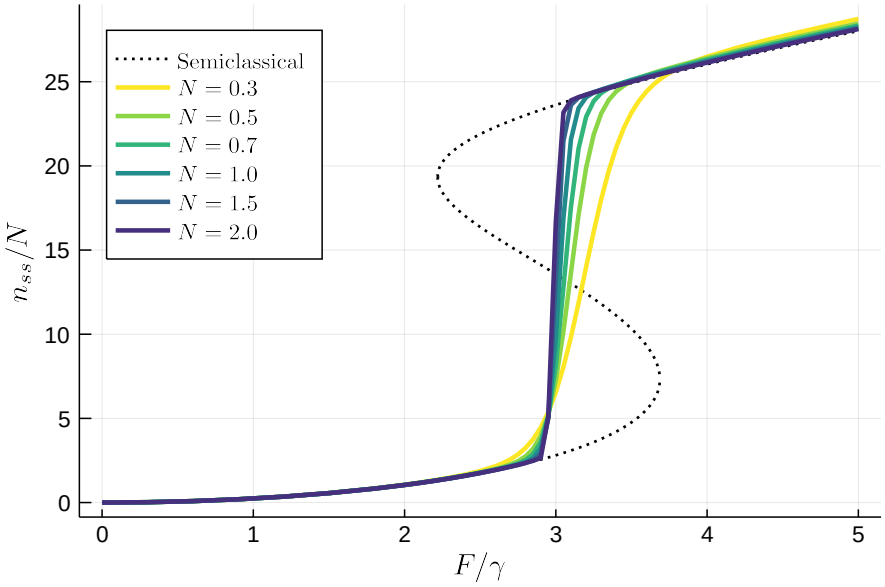

Figure 4.5: Rescaled steady-state expectation value of the occupation number as a function of the pump amplitude F/γ. The dotted line represents the semiclassical solution implying multistability, while the coloured lines are computed for different values of the rescaling factor N. Parameters as in Fig. 4.3: $\Delta = 2\gamma$ and $U = 0.1\gamma$.

A more accurate study has been performed with a self-consistent Gutzwiller mean-field approximation, which approximates the density matrix as

$$\hat{\rho} \approx \otimes_{i=1}^{N} \hat{\rho}_i \tag{4.18}$$

where $\hat{\rho}_i$ is the local density matrix of site i. This approach neglects quantum correlations among sites, but thanks to the self-consistent time evolution it correctly models classical correlations. Unfortunately, also this method does not take into account the spatial dimensionality of the system. Through this approximation Le Boitè et al. have shown that the system still exhibits bistability, (and in some cases even multistability) [214] though the boundaries are slightly shifted compared to the semiclassical mean-field prediction.

4.2.4 Momentum space description

An insight in how the physics of the single-cavity model is related to the physics of the lattice model is obtained by rewriting the driven-dissipative Bose-Hubbard model in momentum space [183]. We perform the canonical substitution $\hat{a}_{\vec{j}} = 1/\sqrt{N} \sum_{\vec{k}} e^{i\vec{k}\vec{j}} \hat{a}_{\vec{k}}$, where N is the total number of sites in the lattice. With some algebra, it is possible to show that linear dissipation in real space transforms into linear dissipation in momentum space, namely,

$$\sum_i \mathcal{D}(\hat{a}_{\vec{j}}) \to \sum_{\vec{k}} \mathcal{D}(\hat{a}_{\vec{k}}). \tag{4.19}$$

The Hamiltonian, instead, can be rewritten as

$$H = \sum_k -(\Delta + 2zJ\cos(k))\hat{a}_k^\dagger \hat{a}_k + \frac{U}{2N}\sum_{k,p,q} \hat{a}_k^\dagger \hat{a}_p^\dagger \hat{a}_q \hat{a}_{q-k-p} + F\sqrt{N}(\hat{a}_0^\dagger + \hat{a}_0), \tag{4.20}$$

where we see that the homogeneous pumping translates in $k-$space to a pump acting only on the $k = 0$ mode. As all modes have the same loss rate, but only one is pumped, modes with $k \neq 0$ will relax to the vacuum unless they are coupled to the homogeneous mode. From Eq. (4.20) it is evident that the coupling arises from the interaction U/N and conserves total momentum. Upon inspection, we see that the coupling term can couple the following modes: (i) the $k = 0$ mode with itself, through the term $\hat{a}_0^\dagger \hat{a}_0^\dagger \hat{a}_0 \hat{a}_0$; (ii) the $k = 0$ mode with 2 other modes $\hat{a}_k^\dagger \hat{a}_{-k}^\dagger \hat{a}_0 \hat{a}_0 + c.c.$; (iii) the $k = 0$ mode with 3 other modes $\hat{a}_k^\dagger \hat{a}_p^\dagger \hat{a}_q - p - k\hat{a}_0 + h.c.$; (iv) four $k \neq 0$ modes. As the action of the ladder operators is $\hat{a}_{k\neq 0}\,|n_k\rangle = \sqrt{n_k}\,|n_k - 1\rangle$, assuming that $\sqrt{n_k} \approx 0$ because the modes are not pumped, we could as a first approximation neglect those terms and study the effective single-mode Hamiltonian

$$H_{k=0} = -(\Delta + 2zJ)\hat{n}_0 + \frac{U}{2N}\hat{a}_0^\dagger \hat{a}_0^\dagger \hat{a}_0 \hat{a}_0 + F\sqrt{N}(\hat{a}_0^\dagger + \hat{a}_0). \tag{4.21}$$

The approximation could be improved by treating the most relevant couplings to the other modes. This effective Hamiltonian is equivalent to that of the Kerr resonator, where the single-particle energy $\Delta \to \Delta + 2zJ$ is rescaled by the lattice hopping, and the interactions and pump amplitude are rescaled by the number of sites of the lattice. Recently, it was suggested by Casteels et al. to study the $N \to +\infty$ of the Kerr model as a proxy of the thermodynamic limit of the lattice model [183]. In Fig. 4.5 we show the rescaled steady-state occupation number $\langle \hat{n} \rangle / N$ of the Kerr model in the $N \to \infty$ limit, which can be interpreted as the mean occupation of the lattice model. The appearance of a discontinuity in the limit $N \to \infty$, suggests that the lattice model might support a first order phase transition in the thermodynamic limit of the bistable region.

4.3 Phase transition in the homogeneous systems

To investigate this discontinuity of the lattice model within a framework that keeps track of dimensionality, we will present a finite-size scaling analysis of statical and dynamical properties of the system in the bistable region.

4.3.1 The numerical method

Numerically, performing a finite-size scaling of a quantum system will incur in an exponential slowdown due to the increase of the dimensionality of the Hilbert space. This problem could be avoid by using the Corner-Space renormalisation method, but that would prevent us from studying dynamical properties. Alternatively, Matrix Product Operators diminish the impact of the increasing dimensionality, but established techniques only work on 1D lattices and at the time when this research was carried out MPO techniques had yet to be applied to 2D dissipative lattices (later, a proof of concept far from a phase transition

was developed by Kshetrimayum et Al. [66]). A viable alternative are phase-space methods based on stochastic differential equations. While in principle a full treatment of the quantum nature of the system would be possible with Gauge-P representation techniques [115, 116], the literature on the subject is still scarce. For that reason, we considered the Truncated Wigner Approximation presented in Section 3.4.3, which works in regimes of weak interaction, relevant for semiconductor-micropillar devices.

The Lindblad master equation is mapped exactly onto a third-order differential equation for the Wigner quasi-probability, representing the density matrix. In the limit of small U/γ, the third-order derivatives can be neglected so that the differential equation becomes a Fokker-Planck equation [111] for a well defined probability function [112, 26]. This equation can be solved via a Monte Carlo average of the solution of the stochastic differential equations for the complex field $\alpha_j(t)$ describing the bosonic mode in the j-th site, namely,

$$\dot{\alpha_j} = \left[-i(\Delta - U(|\alpha_j|^2 - 1) - \gamma/2)\right]\alpha_j - iJ\sum_{j'}\alpha_{j'} + iF + \sqrt{\gamma/2}\,\chi(t), \tag{4.22}$$

where j' runs over the nearest neighbours of j and $\chi(t)$ is a normalised random complex gaussian noise such that $\langle\chi(t)\chi(t')\rangle = 0$ and $\langle\chi(t)\chi^*(t')\rangle = \delta(t-t')$. Within this formalism, expectation values for symmetrised products of operators [112, 26] are obtained by averaging over different stochastic trajectories through the relation $\left\langle\left\{(\hat{a}_i^\dagger)^n, \hat{a}_j^m\right\}_s\right\rangle = \frac{1}{N_{\text{traj}}}\sum_r(\alpha_{i,r}^*)^n\alpha_{j,r}^m$, where the index r runs over the N_{traj} random trajectories.

To validate our approach on the driven-dissipative Bose-Hubbard lattice, we benchmarked the Truncated Wigner Approximation against numerically exact techniques, such as photon counting trajectories [148] combined with the corner-space renormalisation method [69]. The benchmark presented in the Appendix B.1 shows that for small chains of 4 sites (the biggest we could simulate exactly) the Truncated Wigner Approximation provides numerically accurate results for nonlinearities up to $U/\gamma = 0.5$. For our analysis, we fixed $U/\gamma = 0.1$ in order to ensure numerical accuracy. Moreover, this value is within reach for state of the art semiconductor micropillar lattices.

Anticipating a result that will be discussed later, we will show that the Jacobian of the SDE is ill-conditioned in the vicinity of its metastable solutions. As such, to perform a quantitative analysis of dynamical quantities it is essential to employ a stiff solver. Non-stiff solvers such as Euler-Mayurama or Euler-Heun, which normally ensure strong convergence, will only ensure weak convergence on this particular problem, and dynamical quantities will be skewed and be weakly correlated to the selected time-step [98]. As such, to numerically solve the Eq.(4.22) we used a combination of recently developed stiff solvers[1] [161, 160] implemented in *Julia* [217] within the *DifferentialEquations.jl* package [218].

[1] Initial benchmarking suggested that SKenCarp performs poorly on this particular problem. SOSRA2 and SOSRI2 provided excellent results, both in agreement with each other and with the full quantum solution at small lattice sizes. SOSRA2 proved to be the most efficient, and was selected to produce the entirety of the results presented in the following sections of the chapter.

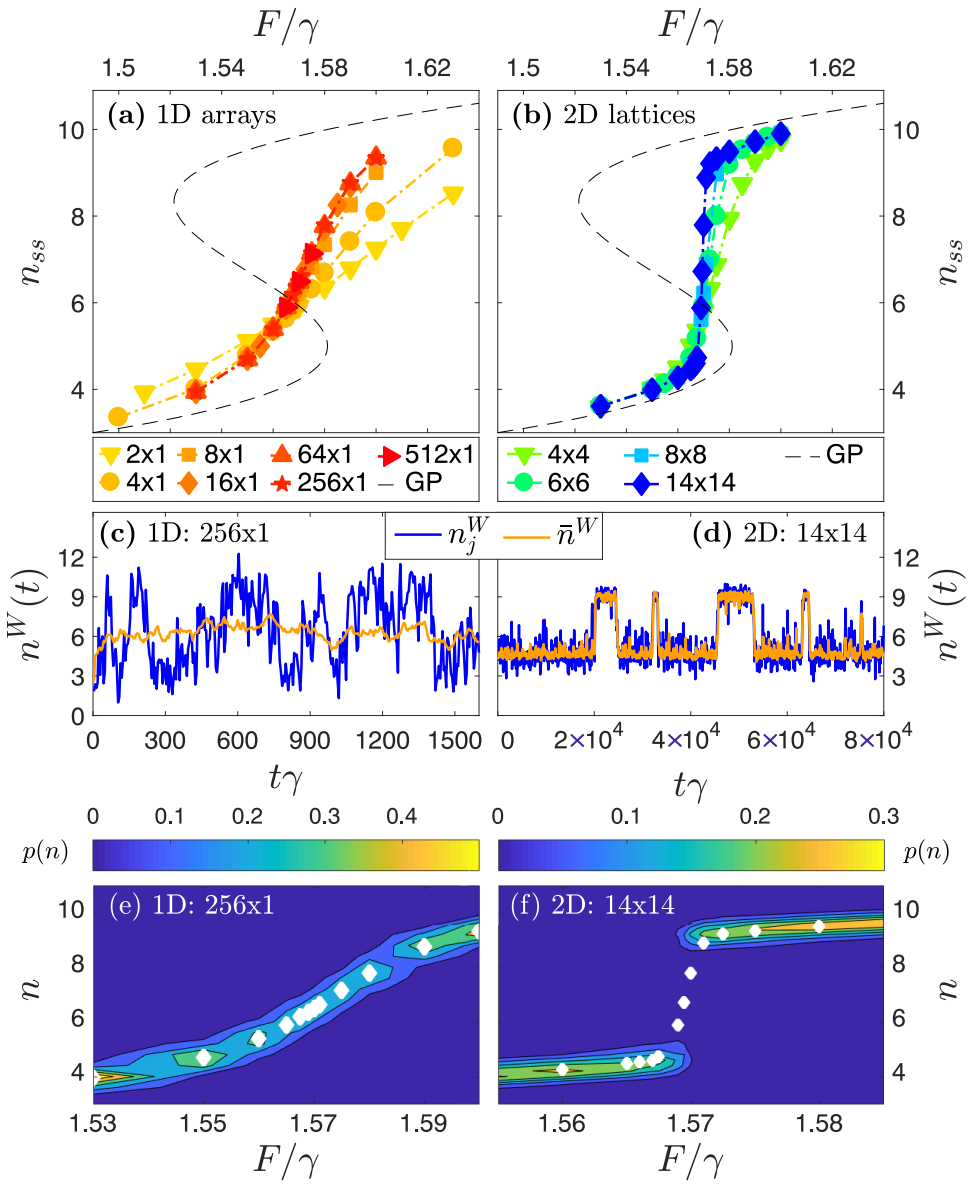

Figure 4.6: Left panels are for 1D arrays, while right panels refer to 2D square lattices. (a) and (b): steady-state average population per site versus driving amplitude F (in units of the dissipation rate γ) for lattices of different size. The dashed line is the prediction of the Gross-Pitaevskii mean-field theory. (c), (d): time-dependent single-trajectory population n_j^W in the $j-$th site (dark blue) and same quantity averaged over all sites $\bar{n}_j^W$ (light orange) for $F = 1.5695\gamma$. (e), (f): contour plot of the probability distribution $p(n)$ of the site-averaged steady-state population versus the driving. White diamonds represent the steady-state average population per site, also shown in panels (a) and (b). (c) and (e) are for a 256×1 array, while (d) and (f) are for a 14×14 lattice. Trajectories have been computed via the truncated Wigner approximation with parameters: $U = 0.1\gamma$, $\Delta = 0.1\gamma$ and $zJ = 0.9\gamma$ (hopping rate times the coordination number z).

4.3.2 Statical and dynamical properties of the lattice boson population

Here we will compare the behaviour of 1D and 2D lattices in the region where mean-field [214, 215] predicts multiple solutions. To make the comparison effective, we will consider the same value of zJ for different lattices, so that the mean field prediction is the same for the 1D and 2D cases, and discrepancies are only due to effects beyond mean-field. In Fig. 4.6(a) we present results obtained with the truncated Wigner approximation for the steady-state site-averaged population $n_{ss} = \frac{1}{N}\sum_{i=1}^{N} \mathrm{Tr}(\hat{\rho}_{ss}\hat{a}_i^\dagger \hat{a}_i)$ for 1D arrays of different length L (up to $L = 512$). In Fig. 4.6(b), the same observable is reported for 2D $L \times L$ lattices (up to 14×14). Both 1D and 2D calculations have been performed with periodic boundary conditions. In both Fig. 4.6(a) and (b) the Gross-Pitaevskii-like mean-field prediction is depicted by the dashed line. Similar to what was seen in our treatment of the Kerr resonator, the multistability predicted by mean-field theories disappears when we target the steady-state density matrix, as the solution, in absence of strong symmetries, is unique under quite general assumptions [52, 37]. In the same way, quantum fluctuations perturb the mean-field solutions so that on a single trajectory the system switches back and forth from one metastable state to another on a time scale related to some eigenvalue of the Liouvillian [112, 181, 219, 197] (see also Fig. 4.6(c)). The results in Fig. 4.6(a) show that the S-shaped multivalued curve of the mean-field theory is replaced by a single-valued function, which depends on the array size L. Remarkably, by increasing the size L of the array $n_{ss}(F)$ eventually converges to a curve with a finite slope. On the other hand, in 2D the slope of $n_{ss}(F)$ does not saturate when increasing the size L of the lattices, suggesting the emergence of a discontinuous jump in the thermodynamic limit compatible with a first-order phase transition.

In Fig. 4.6(c) and Fig. 4.6(d), we present the dynamics of the boson population in a single stochastic Wigner trajectory for the 1D and 2D lattices, respectively. While Wigner trajectories *per se* are a mathematical tool that cannot be given a physical interpretation before averaging, their mathematical definition given by Eq. (4.22) is equivalent (except for a small effective rescaling of the single particle energy given by the term $\propto -U\alpha$) to the dynamics of the semiclassical field. Therefore, Wigner trajectories will show the same qualitative (switching) behaviour of a physical system in the semiclassical regime.

In 1D, switches between the two metastable mean-field solutions are barely visible in the population of the j-th site $n_j^W(t)$ (blue curve) and absent in the site-averaged population $\bar{n}^W(t) = 1/N \sum_{j=1}^{N} n_j^W(t)$ (orange curve), consistent with the formation of moving domains with low and high density inside the array [216]. We remark that domains are not clearly identified in this case because of our choice of parameters, and upon further investigations we understood that by increasing the single particle energy Δ we separate more clearly the low- and high-density branches, which results in more clearly marked domains.

The 2D lattice exhibits a strikingly different behaviour, with random switchings of the average population $n_i^W(t)$ between two well definite metastable states. The populations in all sites switch collectively since $n_j^W(t)$ and $\bar{n}^W(t)$ strongly overlap. Furthermore, notice that the 2D timescales are far longer than in the 1D case, indicating a significantly slower dynamics. The fact that all site populations jump suddenly at the same time suggests that a macroscopic fraction of the population lies in the $k = 0$ mode. This is consistent

with the preliminary analysis presented in Section 4.2.4, where we had claimed, under certain approximations that neglected the effects of dimensionality, that the $k = 0$ would undergo bistable dynamics qualitatively similar to those found in the Kerr single-mode model [183].

Additional insight is provided by the study of the probability distribution $p(n)$ of the occupation number n. Formally, this is defined as

$$p(n) = \frac{\langle n|\hat{\rho}|n\rangle}{\mathrm{Tr}[\hat{\rho}]}, \tag{4.23}$$

and can be obtained with most numerical methods. To compute $p(n)$ at the steady state we select a time $t_s \approx \lambda^{-3}$ where the system on average is steady and statistically collect the values of $n = \bar{n}^W(t > t_s)$ for all the considered trajectories. We plot $p(n)$ in Fig. 4.6(e,f) for different values of the driving amplitude F. We notice that, in the 1D case, this distribution is mono-modal for all values of F and the steady-state mean value of the population follows the peak of this distribution. In the 2D lattice $p(n)$ exhibits a completely different behaviour: it has a single peak in the limit of small and large F, while it is bimodal in proximity of the critical region. Here, for finite-size the steady-state expectation value n_{ss} falls in a region of negligible probability ($p(n_{ss}) \simeq 0$) in-between two peaks corresponding to the low and high population phases. When the 2D lattice size is increased, the crossover between the two phases becomes steeper and therefore the bistable region also becomes narrower, eventually collapsing to a single point when $L \to \infty$. This explains why in large lattices a very fine scan in F is necessary to observe this feature.

4.3.3 Critical slowing down

To investigate the emergence of criticality in the dynamical properties, we calculated the time evolution towards the steady-state value n_{ss} of the site-averaged mean occupation number $n(t) = \frac{1}{N}\sum_{i=1}^{N} \mathrm{Tr}(\hat{\rho}(t)\hat{a}_i^\dagger \hat{a}_i)$, taking the vacuum as initial state. For values of F close to the critical point, $n(t) - n_{ss}$ decays exponentially to zero at large times as reported in Fig. 4.7. As we have shown in Section 2.1.4, in this asymptotic regime the decay dynamics is dominated by the Liouvillian frequency gap λ, which can be extracted by fitting the results with $n(t) = n_{ss} + Ae^{-\lambda t}$. Note that in order to have enough accuracy, calculations have required up to 10^6 stochastic Wigner trajectories for each data point. Experimentally, the asymptotic decay rate can be also measured using the time-dependence of the second-order correlation function [220], dynamical optical hysteresis [197] and switching statistics [192, 197].

The particular case of $F = 1.57\gamma$ is analysed in Fig. 4.7, where we plot $|n_{ss} - n(t)|$ for 1D arrays (panel a) and 2D lattices (panel b) of different sizes. For this fixed value of F, the dynamics gets slower as the size of the simulated system is increased. While in the 1D case the exponential decay rate saturates in the thermodynamic limit, this is not the case for 2D systems. The slowing down of the dynamics is studied quantitatively in Fig. 4.8, where we present the relation between the asymptotic decay rate and the size of the system. The Liouvillian gap λ as a function of the driving amplitude F in 1D arrays is plotted in Fig. 4.8(a): it is apparent that, when the size L is large enough, the Liouvillian gap converges to a finite value for all the values of F, thus proving the absence of critical

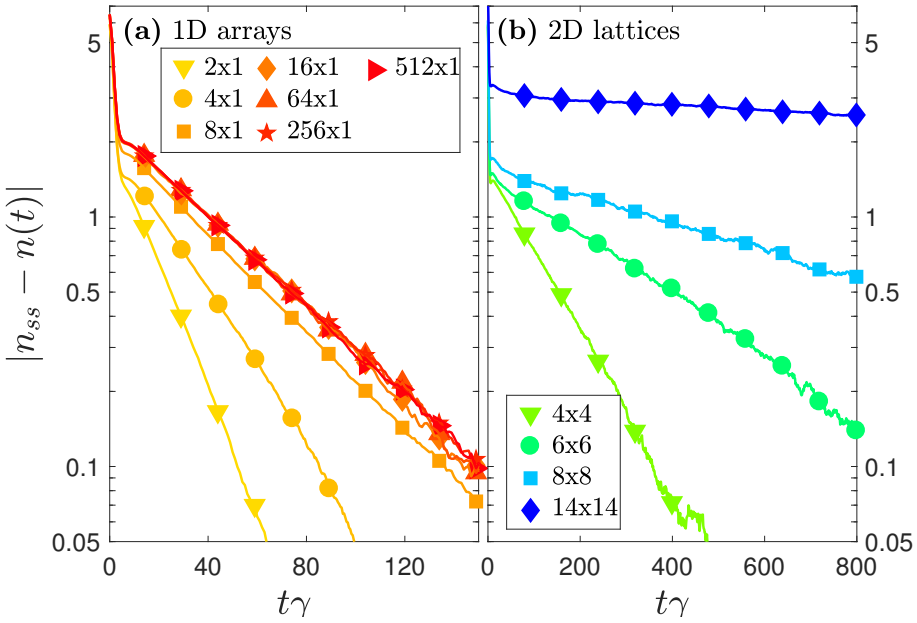

Figure 4.7: Transient dynamics of the absolute difference between the mean occupation number $n(t)$ and its steady-state value n_{ss} for 1D arrays (a) and 2D lattices (b) of different sizes, with driving amplitude $F = 1.57\gamma$. Other parameters as in Fig. 4.6.

slowing down. The behaviour is strikingly different for 2D lattices, as shown in Fig. 4.8(b): in this case, every curve $\lambda(F)$ presents a minimum, which becomes smaller and smaller as the size L of the lattice is increased. As shown in the inset of Fig. 4.8(b), the minimum of λ follows the power-law decay $\min_\lambda(L) \propto L^{-\eta}$, with exponent $\eta = 3.3 \pm 0.1$. Since the phase transition is of first order, this exponent is not universal [183, 220]. To verify this, we computed the critical exponent in lattices with a different nonlinearity (the other parameters were unchanged), finding $\eta = 5.3 \pm 0.1$ for $U/\gamma = 0.06$ and $\eta = 1.7 \pm 0.2$ for $U/\gamma = 0.15$. However, field-theoretic computations suggested that if one where to compute those exponents exactly at the tri-critical point of Fig. 4.3, one would find the same exponents of a thermal Ising model [216, 46].

4.3.4 Steady-state observables

To finally confirm the fact that this phase transition is arising in the homogeneous mode, we studied the fraction of population in the $k = 0$-mode

$$f_0 = \frac{n_{\mathbf{k}=0}}{n_{tot}} = \frac{\sum_{i,j=1}^{N} \langle \hat{a}_i^\dagger \hat{a}_j \rangle}{\sum_{i=1}^{N} \langle \hat{a}_i^\dagger \hat{a}_i \rangle}, \tag{4.24}$$

where $n_{k=0}$ is the steady-state population of the driven $k = 0$-mode and n_{tot} is the total lattice population. In Fig. 4.9(a) and (b) we report the finite-size analysis of f_0 as a function of F. In the region of mean-field bistability, f_0 presents a minimum in both 1D and 2D. In 1D this minimum saturates to a finite value as one approaches the thermodynamic limit, while in 2D f_0 exhibits a behaviour consistent with a finite jump at the critical

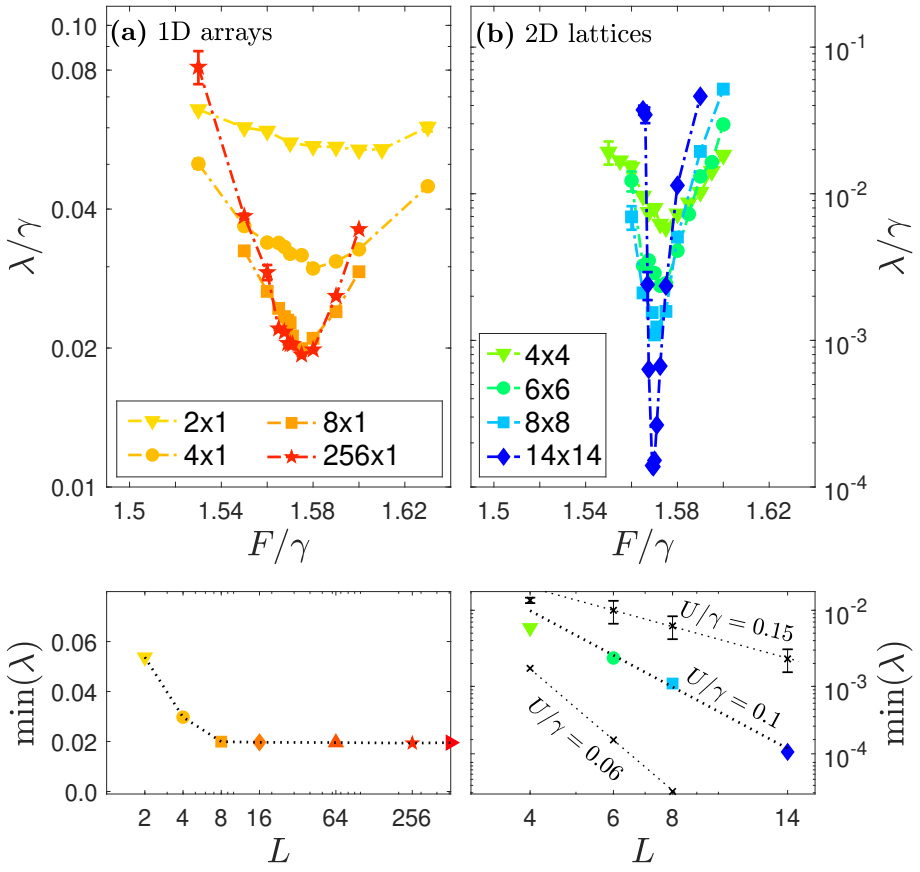

Figure 4.8: The Liouvillian gap λ (log scale) versus the driving amplitude for several $L \times 1$ arrays (a) and $L \times L$ lattices (b). Notice the different scales used for the 1D and 2D case. The bottom panels show the minimum of λ as a function of the size L. For the 2D case we also plot the gap for $U/\gamma = 0.06$ and for $U/\gamma = 0.15$. The dashed line is a linear power-law fit to the data. Error bars are within the symbol size. Parameters as in Fig. 4.6.

point. For the considered interaction, in both cases the population of the driven mode is dominant (f_0 close to 1), showing that the fluctuations induced by the coupling to non-zero momentum modes destroy the critical behaviour in 1D.

To conclude the study of the homogeneous model, we present the local equal-time second-order correlation function $g_0^{(2)} = \langle \hat{a}_j^\dagger \hat{a}_j^\dagger \hat{a}_j \hat{a}_j \rangle / \langle \hat{a}_j^\dagger \hat{a}_j \rangle^2$ as a function of F. This quantity describes the amplitude of the fluctuations in the field, and has been employed extensively to investigate critical behaviour in in optical systems. In 1D this quantity has a broad peak whose shape is shown to converge for large enough L (Fig. 4.9(c)), while in 2D (Fig. 4.9(d)) the finite-size results show an emerging singular behaviour in its derivative at the critical point. The same qualitative behaviour is also observed in the large population limit of a single-mode nonlinear resonator [183, 53], which is equivalent to the $k = 0$ approximation described above.

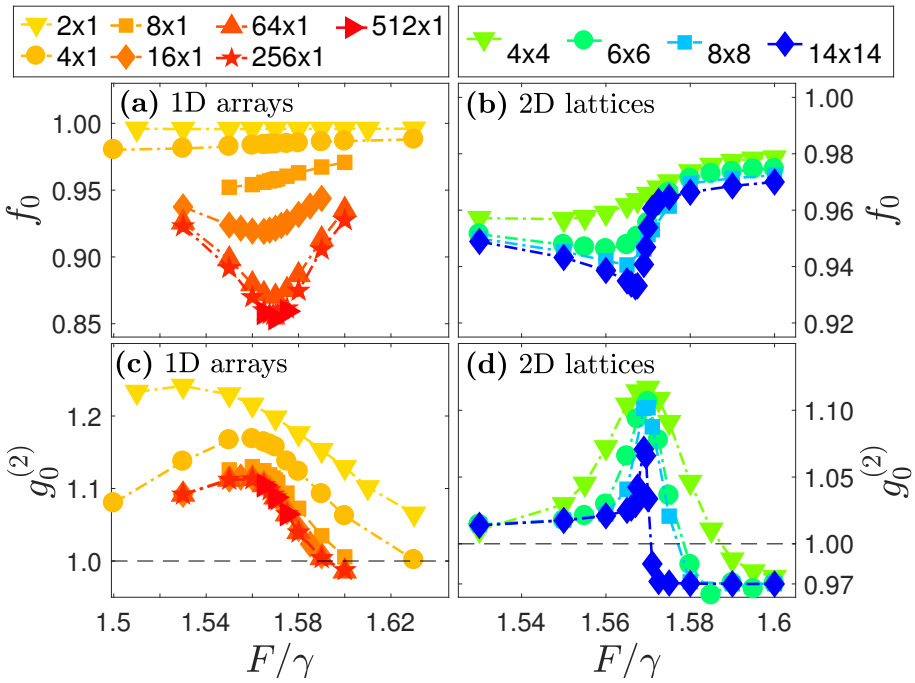

Figure 4.9: (a) and (b): population fraction $f_0 = n_{k=0}/n_{tot}$ in the zero-momentum mode as a function of the driving amplitude. (c) and (d): zero-delay local second-order correlation $g_0^{(2)}$ versus F. Left panels are for 1D arrays, right panels for 2D lattices. Same parameters as in Fig. 4.6.

4.4 Role of disorder on the first-order dissipative phase transition in 2D lattices

In this section we will study how resilient the phase transition that we uncovered in the previous section is, and what happens to the bistability of the $k = 0$ mode. This question is triggered by the fact that previously we have been considering a perfectly homogeneous system where we pumped the homogeneous $k = 0$ mode. Moreover, it is unclear how disorder, which tends to localise excitations and scatter particles into different modes, would affect the criticality.

In what follows, disorder is introduced into the on-site part of the Hamiltonian, by assuming that the local detunings Δ_j are random variables distributed according to a Gaussian distribution $\mathcal{N}(\Delta, W)$ with mean Δ_0 and variance W^2. We will keep all the parameters fixed to the same values as before, and the average detuning $\Delta_0 = \Delta = 0.1$ will match the value used previously.

The calculations presented in this section have been performed combining the Truncated Wigner Approximation (also used in the previous section) with the optimal stochastic unraveling method that we have developed and presented in Section 3.5.3 [2].

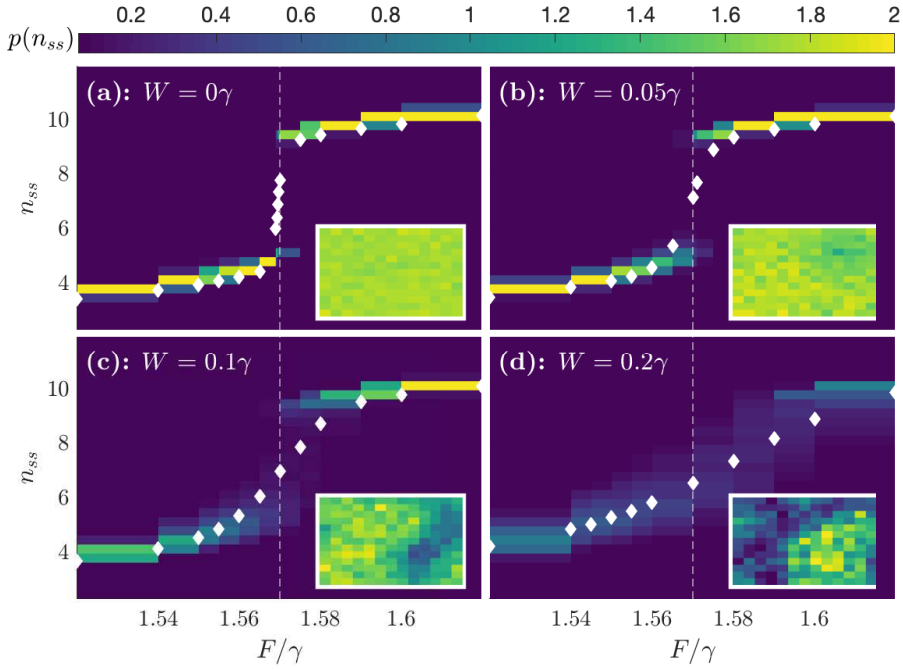

Figure 4.10: Probability distribution of the disorder-averaged steady-state population n_{ss} for different amplitudes W of the disorder. The white diamonds represent the steady-state expectation value $\langle n_{ss} \rangle$. The insets show a typical distribution of the site-dependent boson density at long times for a given disorder configuration. Parameters are $J = 0.225\gamma$, $\Delta = 0.1\gamma$ in a 14×14 lattice. Up to 1000 trajectories, evolved for $t > 4\tau$ were considered for each point.

4.4.1 Effects on the probability distribution of the density

Let us start our analysis by studying the bosonic population across the transition. In Fig. 4.10 we computed the probability distribution of the average steady-state density n_{ss} for several values of the disorder W on a $N = L \times L$ lattice with $L = 14$. In the insets we show typical snapshots of the site-resolved density along a single trajectory, where the formation of domains can be seen for stronger disorder. We note that the abrupt jump present in the clean case [see Fig. 4.10, panel (a)] becomes progressively smoother when influenced by the local disorder. Interestingly, even for disorder distributions well within the single-site linewidth ($W \ll \gamma$) the transition is rounded and replaced by a smooth crossover.

The underlaying mechanism responsible for the suppression of the criticality is the depletion of the homogeneous $\mathbf{k} = 0$ mode of the lattice which is macroscopically occupied across the transition for $W = 0$. In Fig. 4.11, we show the homogeneous fraction f_0 in the steady state for different disorder strengths W. Local inhomogeneities in the energy landscape force the system to populate $\mathbf{k} \neq 0$ modes, competing against the transition. We point out that the strongest depletion takes place for $F < F_{crit}$. Indeed, in the *linear* regime the term that dominates the local energy is proportional to the random

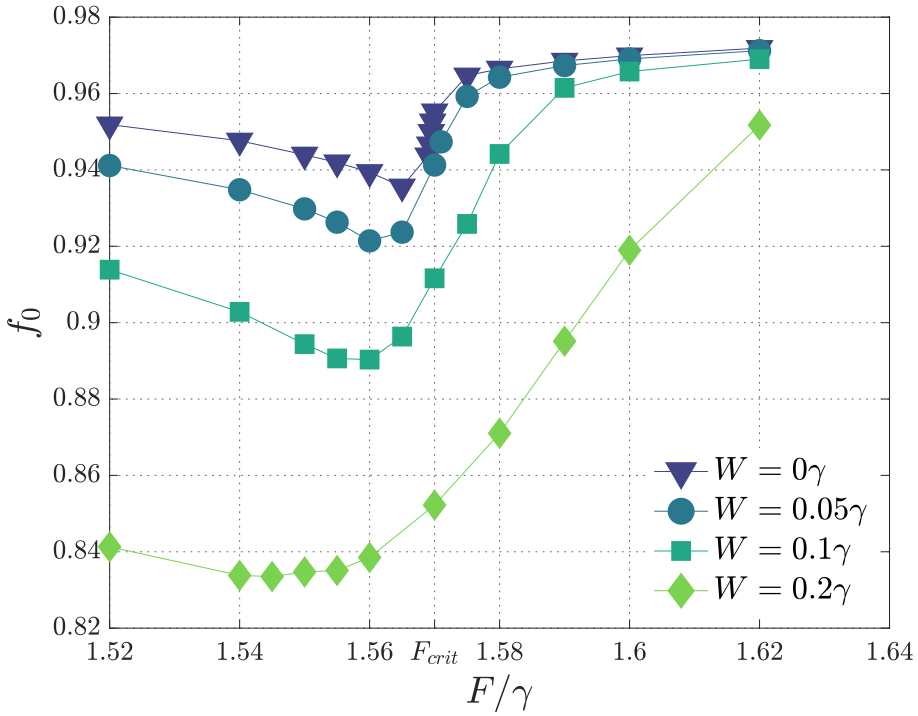

Figure 4.11: Fraction of bosonic population in the $\mathbf{k} = 0$ mode ($f_0 = n_{\mathbf{k}=0}/n$) as a function of F/γ for different disorder amplitudes in a 14×14 lattice. Parameters as in Fig. 4.10.

local detuning. When $F > F_{crit}$ the leading term due to the nonlinearity U grows as n_{ss}^2, making this phase less sensitive to the presence of disorder.

To better characterise the effect of disorder on the criticality, we studied the behaviour of the slope at the transition

$$\mathcal{S} = \left.\frac{\partial n_{ss}(F/\gamma)}{\partial(F/\gamma)}\right|_{F=F^*}, \tag{4.25}$$

where F^* is the value of F for which $\mathcal{S}$ is largest. In Fig. 4.12 we show $\mathcal{S}^{-1}$ as a function of W^{-1} for different system sizes. As the size of the array is increased, the data progressively display a power-law behaviour

$$\mathcal{S} \sim W^{-\beta}, \tag{4.26}$$

with $\beta = 0.98 \pm 0.05$. This result proves that the discontinuous population jump occurring in the thermodynamic limit is smoothened by disorder. Thus only a perfectly homogeneous system ($W = 0$) would display a true criticality $\mathcal{S} \to \infty$ when $L \to \infty$ if one drives the $\mathbf{k} = 0$ mode. This does not exclude that the criticality could be restored by engineering a space-dependent driving field which couples more effectively to the modes of the disordered system.

The value of the exponent β has been extrapolated by fitting the data for $L = 22$. Interestingly, any finite-size system made of $L \times L$ sites, follows the clean-system behaviour in Eq. (4.26) up to a certain value $1/W^*$ which increases as L is increased. This is due to

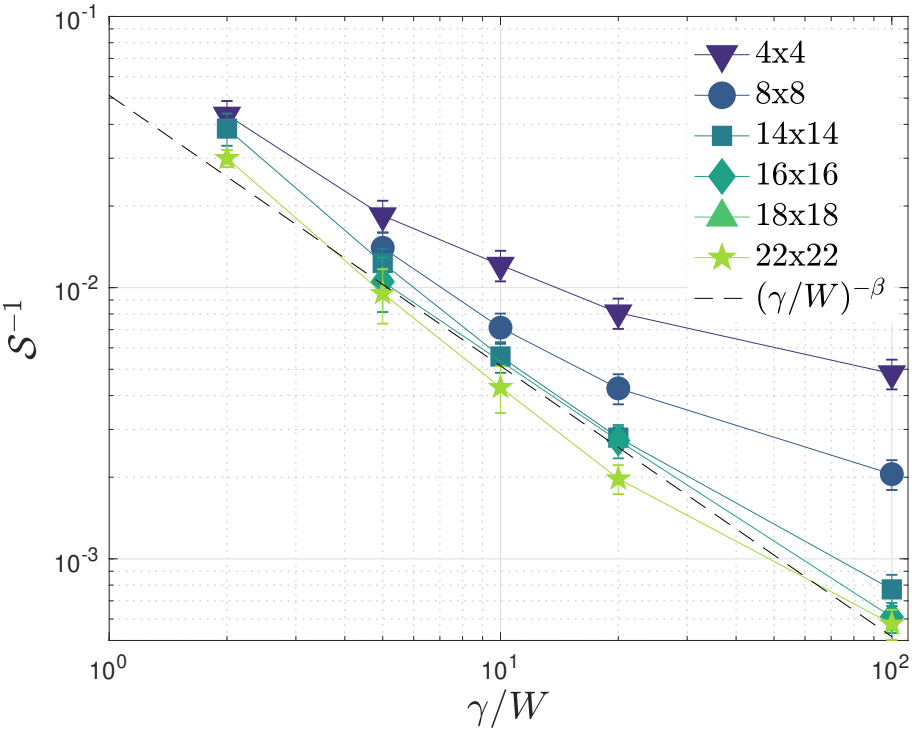

Figure 4.12: Scaling of the inverse slope $\mathcal{S}^{-1}$ as a function of γ/W, for different lattice sizes. The dashed line is the power-law $\mathcal{S}^{-1} = (0.063 \pm 0.010) W^{0.98 \pm 0.05}$ obtained by fitting the 22×22 data. Parameters as in Fig. 4.10.

finite-size effects and, in particular, the relation between W^* and L is connected with the typical correlation length of the systems. To get further insight about this point, in the following section we study the behaviour of the correlation functions at the transition in clean and disordered systems.

4.4.2 Correlation functions

The rounding of the transition due to disorder is also witnessed by the connected part of the one-body correlation functions

$$g^{(1)}(r = |\mathbf{l} - \mathbf{m}|) = \langle \hat{a}_{\mathbf{l}}^{\dagger} \hat{a}_{\mathbf{m}} \rangle - \langle \hat{a}_{\mathbf{l}}^{\dagger} \rangle \langle \hat{a}_{\mathbf{m}} \rangle , \tag{4.27}$$

where $\mathbf{l}$ and $\mathbf{m}$ are the $2D$ coordinates of lattice sites. In a coherently-driven system it is necessary to consider the connected part of the correlators in order not to account for the coherence imprinted by the external driving field [2].

Results are shown in Fig. 4.13 [3]. In the clean case ($W = 0$), the system displays

[2] Given that $\langle \hat{a}_{\mathbf{i}} \rangle \neq 0$, $\forall \mathbf{i}$ because of the local driving term one would simply get $\lim_{|\mathbf{l}-\mathbf{m}| \to \infty} \langle \hat{a}_{\mathbf{l}}^{\dagger} \hat{a}_{\mathbf{m}} \rangle \sim \langle \hat{a}_{\mathbf{l}}^{\dagger} \rangle \langle \hat{a}_{\mathbf{m}} \rangle \to$ const.

[3] Each point is obtained considering the mean value of the correlator Eq.(4.27) averaged over all the pairs of sites at a distance r

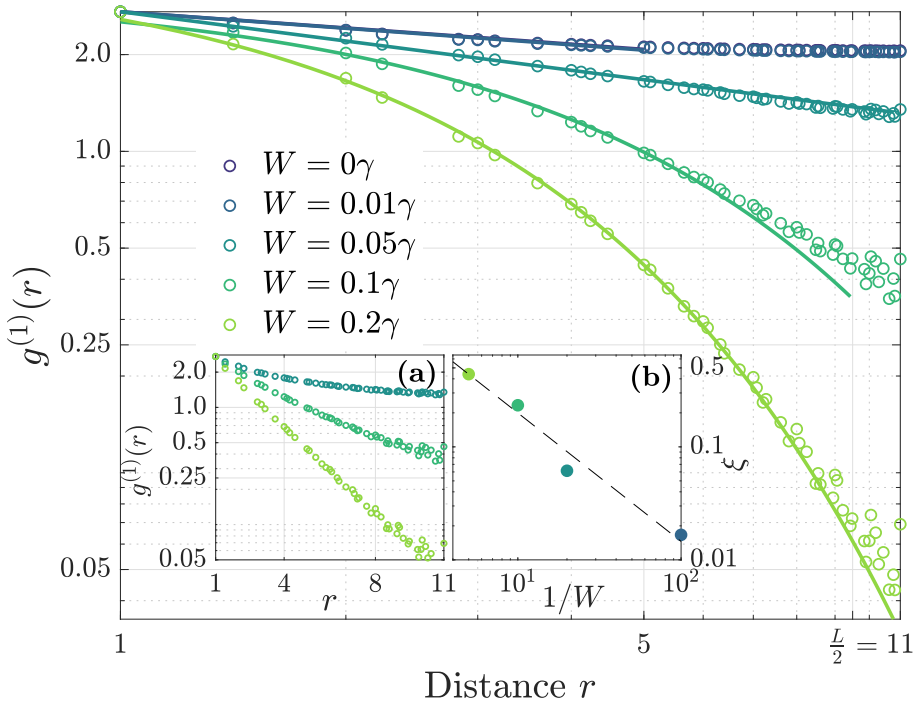

Figure 4.13: Main panel: log-log plot of the two-point correlation function $g^{(1)}(r)$ in a 22×22 lattice with periodic boundary conditions at $F = 1.569\gamma \approx F_{crit}$ for increasing disorder strength W. The dots are the numerical data while the solid lines are the power-law (for $W = 0$) and exponential fits (for $W > 0$). The other parameters are as in Fig. 4.10. Inset (a): semilog plot of the same data in the main panel. Inset (b): log-log plot of the extrapolated correlation length $\lambda(W)$ for $W > 0$. The dashed line is the fit $\lambda(W) \propto W^{-\eta}$ with $\eta = 1.1 \pm 0.2$.

long-range order and the correlation function decays as a power-law

$$g^{(1)}(r) \sim r^{-\alpha}. \tag{4.28}$$

The exponent $\alpha = 0.16 \pm 0.01$ is obtained by fitting the data for $L = 22$. This result is somehow unexpected since in proximity of a first-order transition one would not expect a divergent correlation length [178]. However, this is a peculiar feature of a perfectly clean system and disorder suppresses long-range order.

Indeed, in a disordered system ($W \neq 0$), the correlation function (4.27) decays exponentially as

$$g^{(1)}(r) \sim \mathrm{e}^{-r/\lambda(W)}, \tag{4.29}$$

where $\lambda(W) \sim W^{-\eta}$ (with $\eta = 1.1 \pm 0.2$) is the correlation length which decreases as W is increased. This is a consequence of the formation of coherent domains with typical size $\lambda(W)$. This means that as long as one considers arrays with $L \lesssim \lambda(W)$, the system will behave cooperatively. Indeed, for a system of size L, we do expect departure from the critical thermodynamic behaviour when $L > \lambda(W) \sim W^{-\eta}$. This is consistent to what has been shown in Fig. 4.12, since we find that $1/W^* \sim L^{1/\eta}$.

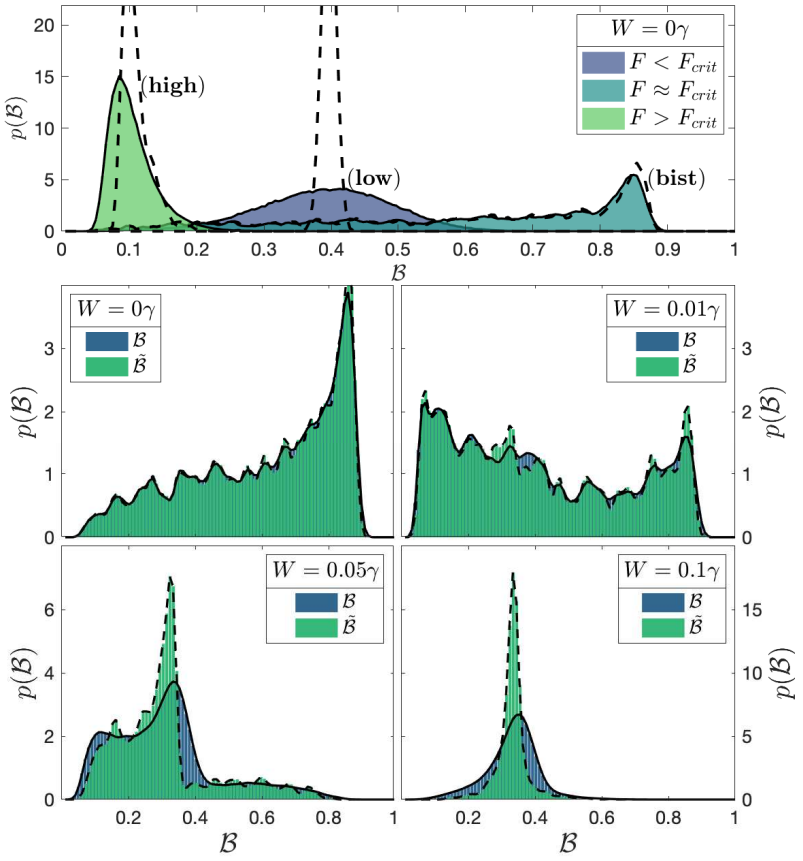

Figure 4.14: Top panel: $p(\mathcal{B})$ and $p(\tilde{\mathcal{B}})$ distribution for the clean case ($W = 0$) for $F < F_{crit}$, $F > F_{crit}$, and $F \approx F_{crit}$, as indicated in the legend. Bottom panel: probability distributions for different values of the disorder amplitude W at $F \approx F_{crit}$.

4.4.3 Collective bistability

In the clean case the phase transition can be also witnessed by studying the dynamics of Wigner trajectories. When $F = F_{crit}$ the lattice exhibits a collective bistable behaviour, dynamically switching between the low- and the high-density phases with a rate linked to the Liouvillian gap. This is due to the fact that the density matrix at the transition is perfectly bimodal [40] and the trajectories will unravel it exploring the two phases with equal probability. Consequently, the probability distribution of the observable witnessing the transition (in this case the boson density) is perfectly bimodal (i.e, it exhibits two peaks with the same probability) at the transition.

To quantitatively characterise the bistable nature of the system near criticality we computed the bimodality coefficient of the density distribution [221]. To this aim, let us define n_i^j as the probability distribution of the density in the i-th site along the j-th trajectory. The bimodality coefficient is then defined as the ratio between the square of

the second central moment and the fourth moment of the distribution [4]

$$\mathcal{B}_i^j = \frac{\mu_2[n_i^j]^2}{\mu_4[n_i^j]}, \tag{4.30}$$

where μ_n is the n-th moment of the distribution. If $n_i^j = \mathcal{N}(n_0, \sigma^2)$ satisfies a Gaussian (monomodal) distribution, then $\mathcal{B} = 1/3$, regardless of its mean value n_0 and variance σ. In the opposite limit, if the distribution is the sum of two normally distributed variables ($n_i^j = [\mathcal{N}(n_A, \sigma_A^2) + \mathcal{N}(n_B, \sigma_B^2)]/2$), then $\mathcal{B} = 1$ for $|n_A - n_B| \gg \sigma_A, \sigma_B$. Therefore these values can be taken as a reference to distinguish between a *perfect* mono- and bi-modal probability distribution. We also define the average bistability coefficient of the j-th trajectory $\tilde{\mathcal{B}}^j = \sum_{i=1}^N B_i^j/N$, by averaging over all the lattice sites for a given trajectory.

Since $\mathcal{B}$ and $\tilde{\mathcal{B}}$ are statistical quantities, they are characterized through their probability distributions. In the following, we will analyze the probability distribution $p(\mathcal{B})$ of the site- and trajectory-resolved bistability coefficient (4.30) comparing it to its site-averaged version $p(\tilde{\mathcal{B}})$. This comparison can be very useful to signal collective synchronized behaviour (such as the bistability expected at the transition) and their fate in the disordered case. Indeed, the average of N uncorrelated identically-distributed random variables is itself a random variable with the same mean and a rescaled variance $\mathbb{V}\mathrm{ar}[\tilde{\mathcal{B}}] = \mathbb{V}\mathrm{ar}[\mathcal{B}]/N$. This is what we do expect in the strong-disorder regime ($W \gg \gamma$). On the contrary, if the dynamical behaviour of different sites is correlated, long-range order forces all resonators to follow the same dynamics. In this case, the covariance among the sites will be maximal, and therefore $\mathcal{B}$ and $\tilde{\mathcal{B}}$ will obey to the same probability distribution.

As a starting point, let us summarize the results for the clean case ($W = 0$) (top panel of Fig.4.14). At criticality ($F = 1.569\gamma \simeq F_{crit}$) the bistable dynamics of the trajectories lead to a distribution $p(\mathcal{B})$ with a peak at $\mathcal{B} \approx 0.85$, which signals the emergence of the bistable collective dynamics. In this case, the fact that the system is behaving cooperatively leads to identical distributions for $\mathcal{B}$ and $\tilde{\mathcal{B}}$. Far from criticality ($F \gtrless F_{crit}$), we obtain that $p(\mathcal{B}) < 0.6$. In particular $p(\mathcal{B})$ is centred at $\mathcal{B} \approx 0.1$ and $\mathcal{B} \approx 0.4$ for $F < F_{crit}$ and $F > F_{crit}$, respectively. This reflects the different nature of the steady-state in the low- and high-density phase. In the first case the $\hat{\rho}^{\mathrm{SS}} \approx |\alpha\rangle\langle\alpha|$ ($|\alpha\rangle$ being a coherent state) leading to quasi-gaussian oscillations around the steady state while in the latter the steady-state is squeezed leading to an asymmetrical non-Gaussian distribution.

We now study the effect of disorder on the bistable behaviour at the transition. In Fig. 4.14 we show the distributions $p(\mathcal{B})$ and $p(\tilde{\mathcal{B}})$ computed for different values of the disorder amplitude W and $F = F_{crit} = 1.57\gamma$. Already for very weak disorder $W = 0.01\gamma$, the clean-system picture is modified. The peak at $\mathcal{B} \approx 0.85$ decreases and the probability to obtain lower values of $\mathcal{B}$ becomes more relevant. However, $p(\mathcal{B})$ and $p(\tilde{\mathcal{B}})$ remain almost superimposed. This means that for very weak disorder the system still behaves as a whole [since $L < \lambda(W)$] even if the number of bimodal trajectories decreases. Increasing further the disorder to $W = 0.05\gamma$ we get a distribution which can be roughly subdivided into three regions: approximately 15% of the trajectories lie in the region with $\mathcal{B} > 0.45$, where

[4] The time-correlation of $\langle n \rangle(t)$ complicates the distribution of B, therefore to decorrelate it we perform a step of gaussian filtering on the temporal data and then we subsample. This is similar to the debunching typically done in Monte Carlo calculations.

$p(\mathcal{B}) \approx p(\tilde{\mathcal{B}})$, meaning that this fraction of randomly generated disorder configurations are still critical and exhibit a bistable behaviour. The rest of the trajectories do not display bistability and can be divided between two peaks, one around $\mathcal{B} \approx 1/3$ and one at $\mathcal{B} \approx 0.1$, as in the clean case for $F < F_{crit}$ and $F > F_{crit}$, respectively. In this range $p(\mathcal{B}) \neq p(\bar{\mathcal{B}})$ since the dynamics of different sites is uncorrelated. Increasing the disorder even further to $W = \Delta = 0.1\gamma$ yields a Gaussian distribution centred around 1/3. In this regime, correlations among different sites are very weak since $L \gg \lambda(W)$ and each site experiences a monomodal Gaussian dynamics uncorrelated with the behaviour of the other sites.

4.5 Conclusions and perspectives

In conclusion, in this chapter we have studied the physics of the driven-dissipative Bose-Hubbard model, which describes for example driven photonic lattices. First, we have shown that the average collective dynamics of an homogeneous lattice can be related to a particular limit of the single-resonator model, where the mapping is expected to be exact in the mean-field limit. Then, we have theoretically predicted the critical slowing down associated with a first-order dissipative phase transition of the model. Through a finite size analysis we have shown that this transition emerges in 2D lattices where it is linked to a finite size bistability of the whole lattice. We have also shown that those features are absent in 1D arrays, where fluctuations destroy criticality of the dynamics even if the driven mode is macroscopically occupied. This suggests that the lower critical dimension must be at least two.

In the following, we have explored the role of disorder in the same dissipative phase transition, revealing that criticality would be suppressed by any amount of finite disorder. We have also exploited our optimal unraveling method to show that correlation functions in the homogeneous system decay algebraically in space. As disorder introduces a natural correlation length in the system, if such length is smaller than the system size it is possible to see a crossover to exponential decays of correlations. By studying the statistical distribution of the time-evolving occupation number along every trajectory we have been able to show that the lattice-level bistability disappears (on average) as soon as disorder is introduced in the system.

Even though we derived those properties for a particular system, we expect the results to be general across the wide class of systems such as the transverse-field dissipative Ising model and the dissipative XY model, which exhibit similar properties, such as bistability and slowing down of the dynamics. Interestingly, relating first order criticalities to lattice-level bistability which only occur when correlation lengths diverge may suggest that those dissipative phase transitions have diverging correlation lengths, which may seem counter-intuitive as in closed-systems the two are unrelated. We believe that this possibility warrants further research.

Additionally, our study is consistent with field-theoretical calculations that suggested that the driven-dissipative Bose-Hubbard model in the tri-critical point recovers in the thermodynamical limit the critical exponents of the classical Ising model [216]. It would be interesting to further investigate how that picture is modified when the transition is superimposed with another quantum phase transition [46].

The emerging picture is directly relevant for the ongoing experiments in state-of-art

photonic quantum simulators based on superconducting circuits and semiconductor microcavities. In these systems, the presence of on-site disorder in the local frequency of resonators is unavoidable. Experimentally, by tuning the size (and/or the disorder) of a sample, one can explore different regimes where collective critical dynamics or density domains dominates the physics.s Lastly, it would be also interesting to study the impact of disorder on a second-order dissipative criticality and to study models where long-range interactions competes with local disorder [222].

CHAPTER 5

Neural Networks for Driven-Dissipative Quantum Systems

In this chapter we introduce an original numerical algorithm to compute the steady state of open quantum systems based on neural networks. To begin, in Section 5.1 we briefly introduce machine learning in a perspective relevant to our work. In Section 5.2 we give a biased introduction to neural networks, giving particular importance to their properties of function approximators. In particular in Section 5.2.3 we introduce energy based models such as Restricted Boltzmann Machines (RBM) and summarise analytical results about their representation power. In Section 5.2.1 we introduce the more general Feed Forward Neural Networks (FFNN) and some conjectures on their representation power. In Section 5.4 we introduce the variational principle for open quantum systems used in our work and compare it with other results on variational methods for open quantum systems. In Section 5.5 we review a first order and a second order method to solve the optimization problem introduced in the previous chapter, and benchmark our method in Section 5.6. Finally, in Section 5.7 we present some unpublished developments to further generalise the network structure by sampling in a constrained purified space.

Contents

5.1 Overview of machine learning and neural networks

5.1.1 Historical introduction

Since the dawn of the human kind, persons have understood that they are not efficient at performing certain tasks, and have built tools and devices that are more adequate. Since the first programmable computer was built, scientists have identified a set of problems that are hard to solve for humans, but which are relatively straightforward for a computer, such as matrix calculus. Those problems are usually described and solved by a finite list of formal, mathematical rules. In the field of *artificial intelligence* scientist are interested by another set of tasks, where humans excel but machines perform relatively weakly, such as image recognition. Those problems are solved intuitively by humans, and they cannot be easily phrased in terms of mathematical rules.

Machine learning

Machine learning is a scientific field that studies how to program computers to efficiently solve artificial intelligence tasks. Instead of explicitly identifying a set of rules, we let the computers learn from experience and identify by themselves the rules to solve the task. This is achieved by specifying only the desired high-level behaviour of the program (solution) and its high level structure. In more mathematical terms, we are not selecting a single program (by programming it) in the *program-space*, but we are rather defining a whole subset of programs, and ask the computer to identify the one that better performs in this set. From this point of view, machine learning rephrases the problem of writing specific algorithms to optimising a generic algorithm. In general, a program solving a certain task is a function that takes some input and returns an output. Therefore in the rest of the chapter we will speak about optimising functions[1].

Extensive research in the field of machine learning revolves around how to parametrise the program-space which contains the possible solutions to the task or problem considered. Answers to the question date back to the late 50s, when Rosenblatt invented the *perceptron* [224], an algorithm for performing pattern recognition. Rosenblatt had essentially rephrased in mathematical language the biological behaviour of a neuron.

[1] Some people refer to this *new* way to think software programming as Software 2.0 [223].

The idea of combining several such neurons into a dense network to describe a more complex function, which can be tuned by changing internal parameters, had already been considered in the late 60's [225] but practical issues delayed the surge in popularity to the new millennia. The overarching problem was that computers were not powerful enough to perform the kind of calculations necessary to optimise those networks. Moreover, efficient optimisation algorithms for those kind of problems did not exist and were only proposed more than 15 years later [226]. Thanks in part to Moore's law and in part to the development of specialized hardware, computers today can perform roughly 10^9 more operations per second than in the $70's$ [227], and research in the field restarted.

There are essentially two questions that people try to answer in machine learning:

- How to write general solutions to some class of problems, which must later be optimised;
- How to optimise those general solutions.

In a language more familiar to physicists, the first question is the problem of finding a good yet generic variational ansatz, while the second is how to optimise the ansatz.

Supervised and unsupervised learning

Machine learning algorithms are techniques to optimise a general solution $f \in \mathcal{H}$ in the space of all solutions $\mathcal{H}$ and find the best-guess. Algorithms can be distinguished into two broad categories, **supervised** and **unsupervised** learning, distinguished by what kind of dataset they are allowed to consume during the learning process.

- **Supervised** learning algorithms consume a labelled dataset $\{(\mathbf{x}, \mathbf{y})\}$, which means that to every entry $\mathbf{x}$ in the dataset corresponds some target label y that we wish to reconstruct. Roughly speaking, curve-fitting is some sort of (very trivial) supervised learning, where given a set of points $(\vec{x}, y)$ the objective is finding the $f(\mathbf{x})$ that best minimises the distance χ^2.
- **Unsupervised** learning algorithms consume a datasets $\{\mathbf{x}\}$, which is usually a set of samples taken from an unknown distribution $p(\mathbf{x})$, and attempts to find the approximated distribution $f(\mathbf{x})$ which best reproduces $p(\mathbf{x})$. This can be achieved, for example, by minimising the Kullback-Leibler divergence between the two distributions $KL(f, p)$, which plays the same role of the χ^2 in the previous example [228].

In the two examples above, the solution is found by minimising the χ^2 and the K-L divergence respectively. Those functions are known as *cost functions*, and play a similar role to the energy-functional that must be minimised to find the best solution in physics. We remark that cost functions usually do not carry much information about the solution to the problem, but are merely computationally cheap quantities that can be optimised to improve the solution f. The main difference between machine learning techniques and more traditional optimisation is that the first avoids performing the optimisation exactly, resorting to approximated algorithms. It is therefore crucial that the cost function is some quantity that can be computed approximately cheaply.

Neural Networks

Regardless of the optimisation algorithm used, one must still define the subset $\mathcal{H}$ of the function space where to search for solutions. Historically, people experimented with a certain class of functions named *neural networks* and witnessed that they perform well in encoding those solutions. Neural networks are parametrised functions, loosely inspired by how a real neuron works, which are often used in the context of machine learning because of their versatility.

The power of networks arises from chaining several nonlinear functions together, which allow for a more dense encoding compared to an expansion on a linear basis. While at first research focused on very small neural networks, with only one nonlinear function, it has later been understood that the optimal performance comes when several such functions are chained into a structure called *deep* neural network. Interestingly, Kolmogorov and Arnold had already noticed this property in the latter part of the 1950s, and published few papers where they proved that a chaining two sets of particular nonlinear functions could be used to approximate any other smooth function with a polynomially bounded number of parameters [229, 230].

5.2 Deep feedforward networks as universal approximators

Deep feedforward networks, also called feedforward neural networks (FFNN) or multilayer perceptrons, are the quintessential deep learning model [228]. The goal of a feedforward network is to approximate an arbitrary function $f^\star$. A feedforward network defines a mapping $\boldsymbol{y} = f_{\boldsymbol{\theta}}(\boldsymbol{x})$, where $\boldsymbol{\theta}$ are the parameters of the representation that achieve the approximation $f_{\boldsymbol{\theta}} \approx f^\star$.

These models are called feedforward because information flows from the input vector $\boldsymbol{x}$ to the output $\boldsymbol{y}$ through a chain of intermediate computations. There are no feedback connections in which the output of a computation is fed back into some previous layer. They are called networks or graphs because it is customary to represent the mathematical formulas that define them as a graph. Feedforward networks correspond to directed acyclic graphs, describing how functions are composed together. Usually nodes represent a sum over a vector of inputs and the application of a non-linear function. Links among nodes represent the multiplication of the input by a parameter.

5.2.1 Feedforward neural networks

A feedforward neural network is typically written as a chain of M functions $f_1, \dots, f_M$, as $f(\boldsymbol{x}) = f_M \circ f_{M-1} \circ \cdots \circ f_1(\boldsymbol{x})$. In this chain, f_i is the $i-$th layer, and f_M is the output layer. Every layer is a function $f_i : R^N \to R^M$, and the dimension of the domain is called **width** of the layer. The number of layers gives the depth of the network. Networks with only one non-linear layer are called shallow or single-layer models, whereas the rest are known as deep networks. In a deep network, the domain and codomain of the function that we want to approximate constrain the width of the input and output layer respectively. All the intermediate layers can have arbitrary width. Usually, the output layer of a network is a simple linear function like the sum $\sum$.

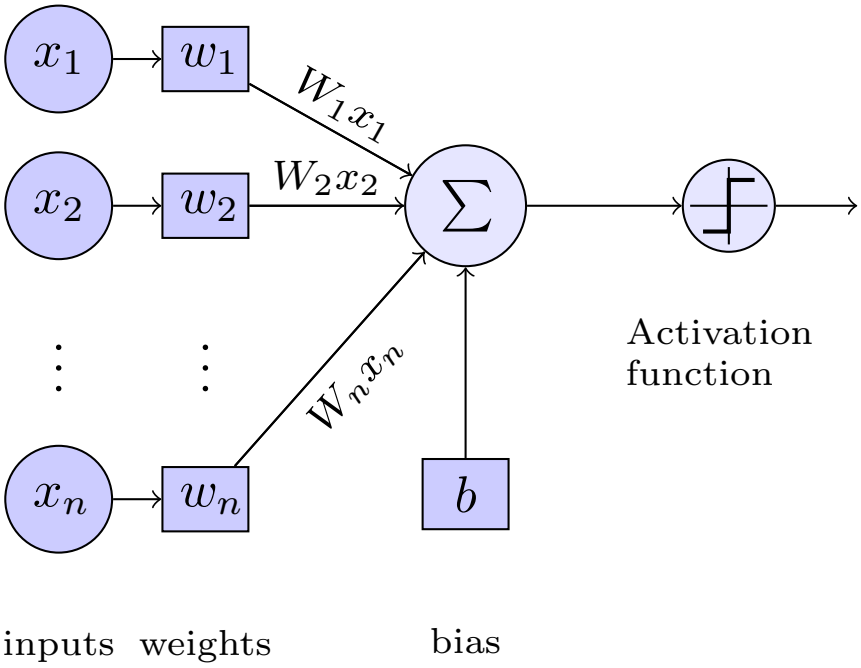

Figure 5.1: Graph representation of single neuron (perceptron) as described by Eq. (5.1). The neuron takes as input a vector of n values $\{x_1, \dots x_n\}$, performs the weighted sum with weights $\{w_1, \dots w_n\}$ and shifts the result by the bias b. The result is then passed through a nonlinear activation function σ.

We call those networks *artificial neural networks* because they were loosely inspired by neuroscience and by how neurons work. Neuroscientists model neurons as a scalar nonlinear function $n : \mathbb{R}^N \to \mathbb{R}$ acting on some linear mixing of the inputs $\boldsymbol{x} \in \mathbb{R}^N$, as depicted in Fig. 5.1,

$$n(\boldsymbol{x}) = \sigma \left[b + \sum_j W_j \, x_j \right], \tag{5.1}$$

where $\boldsymbol{W}$ is a linear mixing vector, b is a scalar bias, and σ is an arbitrary continuous nonlinear bounded function (usually a sinh), called **activation function**.

Layers of a feedforward networks are then built by combining several of those neurons in parallel, generating a function mapping vectors $\boldsymbol{x} \in \mathbb{R}^N$ to vectors in $\mathbb{R}^M$, according to the following function $\boldsymbol{D}(\boldsymbol{x}) = \{\boldsymbol{N}_i(\boldsymbol{x})\}_{i=1\dots M}$ such that

$$\boldsymbol{D}(\boldsymbol{x}) = \boldsymbol{\sigma} \circ [\boldsymbol{b} + W\boldsymbol{x}], \tag{5.2}$$

where $\boldsymbol{\sigma} = \{\sigma_i\}$ $\forall i = 1 \dots M$ is a set of activation functions, the Hadamard symbol $\circ$ denotes element-wise application of a function on a vector, $\boldsymbol{b} = \{b_i\}$ $\forall i = 1 \dots M$ is a set of biases acting like the bias b in a single neuron, and $W \in M(\mathbb{R}^M, \mathbb{R}^N)$ is a linear mixing matrix.

When all σ_i are identical $\boldsymbol{D} : \mathbb{R}^N \to \mathbb{R}^M$ is called a dense layer. The simplest FFNNs are obtained by chaining several dense layers together into $f = \sum \circ \boldsymbol{D}_3 \circ \boldsymbol{D}_2 \circ \boldsymbol{D}_1$.

5.2.2 Representation power of feedforward networks

In general, the representation power of a parametrised function is related to its ability of approximating up to arbitrary precision, generic functions. While a precise definition does not exist, a neural network is said to have higher representation power than another one if it can approximate more functions with the same number of parameters.

The representation power of neural networks is an active domain of research as it allows to better understand their limitations. In the last decades several rigorous theorems on the representability of networks have been proven. To convince even the most sceptic reader that we do understand why neural networks work, we will present a series of results that prove that networks can be used as approximations to arbitrary complex functions.

Theorem 1 (Universal Approximation Theorem)**.** *Let $\mathcal{I}_M$ denote the M-dimensional unit hypercube $[0,1]^M$ and $C(\mathcal{I}_M)$ the space of real-valued continuous functions on $\mathcal{I}_M$. Let $f : \mathbb{R}^M \to \mathbb{R}$ be a single-layer dense neural network $f = \sum \circ \boldsymbol{D}$ with width N and arbitrary activation function σ (continuous, bounded and non-constant) :*

$$f = \sum \circ \boldsymbol{v}\boldsymbol{\sigma} \circ [\boldsymbol{b} + W\boldsymbol{x}]$$

where $W \in \mathbb{R}^N \times \mathbb{R}^M \wedge v, b \in \mathbb{R}^M$. Given any function $f^\star \in C(\mathcal{I}_M)$ and any arbitrary error $\epsilon > 0$, there exist a width N and set of parameters $(W,\ \boldsymbol{v},\ \boldsymbol{b})$ so that

$$|f^\star(x) - f(x)| < \epsilon \quad \forall x \in \mathbb{R}^M$$

Equivalently, single-layer dense neural networks are dense in $C(\mathcal{I}_M)$.

Initially the theorem was formulated only for the sigmoid activation function, $\sigma(x) = 1 + e^{-x}$ [231], but later it was understood that this result is more general and valid for any activation function [232][2]. The theorem essentially proves that in the limit of $\lim_{N\to\infty}$, namely in the limit of an infinite number of parameters, a single-layer FFNN can approximate any continuous function on a compact support. However, the size of such a neural network can be exponential in the input dimension [234]. It has been shown that Artificial Neural Networks (ANN) struggle with highly varying functions, particularly when they have a high density of local extremas. In those cases the single-layer FFNN reduces to performing a piece-wise approximation of the function, requiring a width exponential in the dimension of the input [235, 236].

Even though shallow networks are inefficient encodings of arbitrary functions, deep networks are much more efficient. In 2015 it was proven by Eldan et al. that given an arbitrary function $g : \mathbb{R}^M \to \mathbb{R}$, there exist a two-layer deep dense FFNN with width at most $CM^{19/4}$ approximating to arbitrary precision g. The constant C depends linearly on the precision (see section 2 in [235] for a detailed formulation of the theorem, and section 3 for a proof).

[2]The theorem is actually even more general, as it has been generalised to the ReLU *activation function* $f(x) = \max(0, x)$, which is not continuous nor nonlinear or bounded [233].

Comparison to the expansion on linear basis

Chaining nonlinear functions together is not the only way to approximate functions. Another simple alternative involves writing the function as a linear combination of some basis elements. This expansion is the foundation of the Fourier expansion, but it is also widely used in physics when functions acting on an Hilbert space are expanded in the position basis or the k-space basis. Also Matrix Product States algorithms are based on this expansion. Those methods are based on the fact that the control parameter of the approximation (such as the maximum momentum k in the Fourier expansion, or the bond-link dimension b in an MPS state) can be linked to what regions of the function-space are neglected. It is therefore easy to understand, at least with respect to the artificial neural network approach, what can be represented and what cannot be represented when truncating the space.

On the contrary, the space of (well behaved) Lebesgue-integrable functions on $\mathbb{R}^M$ has dimension $\dim(L^2(\mathbb{R}^M)) \propto e^M$. Those linear expansions are not efficient encodings for arbitrary functions. Deep FeedForward Networks, in principle, are efficient encodings.

5.2.3 Energy based models

In this section we have introduced Deep Feed Forward Neural networks as efficient encodings of arbitrary functions defined on a Real-valued. Other then the constraints set by the representation power, a feedforward network can approximate *any* function. Although, in physics we are often interested in describing particular functions, such as probability distributions, density matrices or a many-body wavefunction. Those functions are not arbitrary, but satisfy certain constraints. To avoid performing constrained optimisation, which is a harder problem than unconstrained optimisation, it is of interest to devise ANN models which satisfy those constraints by construction.

A simple yet useful constrain that can be enforced on the network is positivity of the output, so that upon normalisation it can be considered a valid probability distribution. A simple way to enforce this condition is by exponentiating an arbitrary real-valued function $E(\boldsymbol{x})$, so that $p(\boldsymbol{x}) = \exp[-E(\boldsymbol{x})]$. Since $\exp(z)$ is positive $\forall z \in \mathbb{R}$, this guarantees positivity. To increase the representative power for distributions of the variables $\boldsymbol{x} \in \mathbb{R}_V$, a bigger space $\mathbb{R}_V \otimes \mathbb{R}_H$ is used, where $\mathbb{R}_V$ is the domain of the distribution and $\mathbb{R}_H$ is a latent space used to encode hidden degrees of freedom and correlations among the physical variables.

$$p(\boldsymbol{x}) = \sum_{\boldsymbol{h} \in \mathbb{R}_H} \exp[-E(\boldsymbol{x}, \boldsymbol{h})] \tag{5.3}$$

Models described by an equation of the form Eq. (5.3) are called *energy-based models*, because the function E can be interpreted as an energy functional on the space $\mathbb{R}_V \otimes \mathbb{R}_H$. When the variables $\boldsymbol{x}$ and $\boldsymbol{h}$ are allowed only binary ± 1 values, the energy functional identifies a spin-glass model. For that reason, the study of high-dimensional glassy models is deeply linked with the representation power and learning of neural networks.

A paradigmatic example of such a model is the *Restricted Bolzmann Machine* (RBM) depicted in Fig. 5.2, which is based on the energy of a spin-glass with bipartite all-to-all connections. Let us suppose that input variables are binary spins (this can be generalised to higher dimensional spins without major difficulties). Then, a generic state of the system

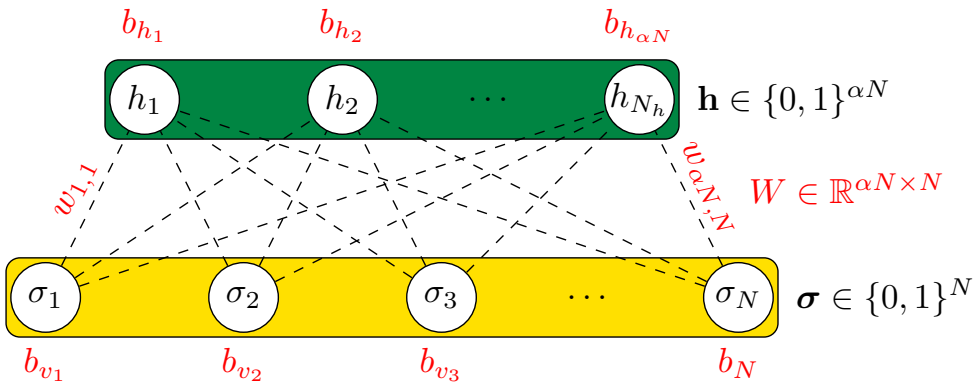

Figure 5.2: Graph representation of the energy-based model defined by Eq. (5.4). Spins are divided into two partitions, the lower (visible) lives in the same space as the target distribution, while the top (hidden) is a latent space used to encode information. α is the hidden space density. Interactions are between all spins of one partition with all spins of the other partition.

is $(\boldsymbol{\sigma}, \boldsymbol{h})$ where $vb * \sigma$ is part of the *visible layer* and $\boldsymbol{h}$ is part of the *hidden layer*. The energy functional of a RBM reads

$$E(\boldsymbol{\sigma}, \boldsymbol{h}) = \boldsymbol{\sigma}^T \cdot \boldsymbol{b}^{(v)} + \boldsymbol{h}^T \cdot \boldsymbol{b}^{(h)} + \boldsymbol{\sigma}^T W \boldsymbol{h}, \tag{5.4}$$

where $\boldsymbol{b}^{(v)}$ is the local energy term (bias) for the visible layer, $\boldsymbol{b}^{(h)}$ is the bias for the hidden layer and W is the matrix describing the interactions.

We wish to represent probability distributions $p(\boldsymbol{\sigma}) : \mathbb{R}^M \to \mathbb{R}$ defined on the space $\mathbb{R}^M$. Setting the visible layer so that $\boldsymbol{\sigma} \in \mathbb{R}^M$ and replacing Eq. (5.4) in Section 5.2.1 and taking a partial trace over the hidden layer, we obtain

$$p(\boldsymbol{\sigma}) = \sum_{\boldsymbol{h}} \exp[-E(\boldsymbol{\sigma}, \boldsymbol{h})] = \exp\left[-\boldsymbol{\sigma}^T \cdot \boldsymbol{b}^{(v)}\right] \prod_{j=1,\dots,N} 2\cosh\left(b_j^{(h)} + \boldsymbol{W}^{(j)}\boldsymbol{\sigma}\right). \tag{5.5}$$

The restriction to only bipartite interactions among the spins is what allows to carry out analytically the summation over the latent space, and ultimately to obtain those analytical formulas. This greatly reduces the computational cost of evaluating restricted Boltzmann machines, which would otherwise require performing the trace summation numerically.

We remark that the log-probability $\log[p(\boldsymbol{\sigma})]$ encoded in a restricted Boltzmann machine can be written as a feedforward network with a dense layer if the on site terms $\mathbf{b}^{(v)}$ are neglected[3]. An example of this representation is given in Fig. 5.3.

[3]It could be written as a feedforward network with a more exotic intermediate layer, but as that bias term does not usually encode much information, canonical applications of RBMs neglect it.

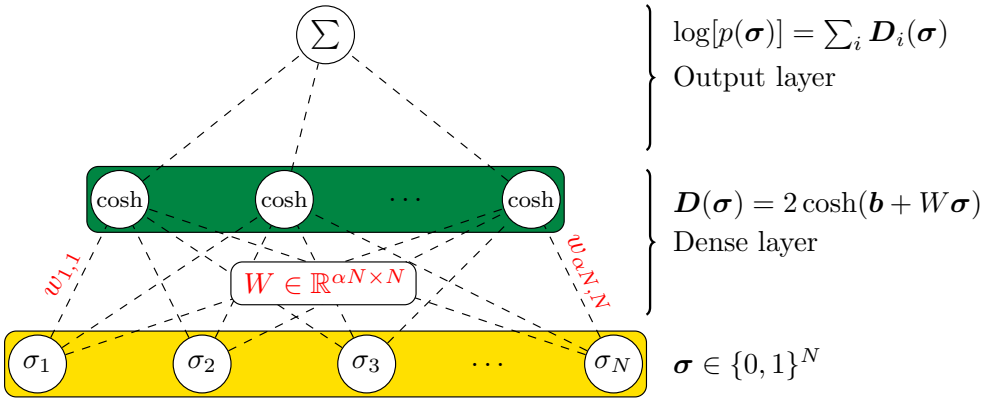

Figure 5.3: Representation of the restricted Bolztmann machine defined by Eq. (5.4) and Fig. 5.2 as a single layer feedforward neural network encoding the log-probability $\log[p(\boldsymbol{\sigma})]$. The on-site energy term $\mathbf{b}^{(v)}$ is not shown here, as it cannot be written as a dense layer. The first and only intermediate layer applies the linear mixing W and the bias $\mathbf{b}$ to the inputs, and then applies the nonlinear function $2\cosh$. The output layer (top part) simply sums over the result of the hidden layer.

Representation power of RBMs

The restricted Boltzmann machine represented by Eq. (5.5) encodes the probabilities in the parameter of the spin-glass energy functional. Namely, if the domain $\boldsymbol{\sigma} \in [0,1]^N$ and $\boldsymbol{h} \in [0,1]^M$, then we are approximating functions on a $N-$dimensional space using an $M-$dimensional latent space. It follows trivially that $\boldsymbol{b}^{(v)} \in \mathbb{R}^N$, $\boldsymbol{b}^{(h)} \in \mathbb{R}^M$ and $W \in \mathbb{R}^N \otimes \mathbb{R}^M$. By increasing the latent-space dimension M, one increases the number of parameters and therefore increases the representation power of this functional.

5.3 Approximating quantum states with neural networks

In the previous section we have shown that neural networks can be used to approximate arbitrary real-valued functions, and can be further constrained to be positive in order to encode probability distributions. In this section we will show how this can be applied to approximate quantum states.

5.3.1 Neural quantum states

For simplicity we will start by considering closed Hamiltonian systems defined on an Hilbert space $\mathcal{H}_S$. The state of said system is encoded into a wavefunction $|\psi\rangle \in \mathcal{H}_S$. If $\{|\boldsymbol{\sigma}\rangle = |\sigma_1, \sigma_2, \ldots, \sigma_N\rangle\}$ denotes a basis of states for the Hilbert space $\mathcal{H}_S$, the wavefunction can be expanded onto this basis according to the definition

$$|\psi\rangle = \sum_{\sigma_1,\sigma_2,\ldots,\sigma_N} \psi(\sigma_1, \sigma_2, \ldots, \sigma_N)\, |\sigma_1, \sigma_2, \ldots, \sigma_N\rangle\,. \tag{5.6}$$

The scalar, complex-valued wavefunction $\psi(\sigma_1, \sigma_2, \ldots, \sigma_N) : \mathcal{H}_S \to \mathcal{C}$ describes the state of the system.

The dimension of the Hilbert space of a quantum system increases exponentially with the number of components. As already discussed in Chapter 3, this growth hampers the possibility of storing the whole wavefunction in vector form for relatively small systems when performing numerical simulations. Therefore it is of paramount importance to find ways to compress or approximate the wavefunction with a lower dimensional encodings. A standard technique involves restricting the wavefunction to be described by a specific parametrised function, known as *variational ansatz*. The ansatz maps a set of parameters $\boldsymbol{v}$ into a wavefunction $\psi_{\boldsymbol{v}}$ defined by

$$\psi : \mathcal{W} \to \mathcal{H}_S, \tag{5.7}$$

$$\boldsymbol{v} \to |\psi_{\boldsymbol{v}}\rangle\,. \tag{5.8}$$

A well-constructed variational ansatz usually has the property $\dim \mathcal{W} \ll \dim \mathcal{H}$, so that it can be stored more efficiently than the whole wavefunction. The price to pay is being limited to only a certain class of wavefunctions $\psi_{\boldsymbol{v}}$. There are many types of ansatz, each with its own pros and cons, and with different regimes of applicability, such as the Gutzwiller ansatz [237], the matrix product states ansatz [238], the projected entangled pair states ansatz for 2D systems [239], the entangled plaquette states (EPS) ansatz [240].

The combination of the variational principle with variational ansätze gave rise to the field of Variational Monte Carlo (VMC) methods [241] which have shown remarkable progress in the last decades [242, 243]. Recently, Carleo and Troyer remarked that, as artificial neural networks are low-dimensional approximations to extremely high-dimensional functions, they are a good fit to be used as variational ansätze for a many-body wavefunction [70]. Restricting ourselves to a system of N spins, Carleo and Troyer proposed to construct a Restricted Boltzmann Machine where physical spins $\sigma_1, \sigma_2, \ldots, \sigma_N$ are identified with the input layer. Even if they introduced the idea with spins, the approach is general and can be generalised to any system with a locally-finite Hilbert space[4]. The resulting ansatz is the following:

$$\psi_{\boldsymbol{v}}(\boldsymbol{\sigma}) = \exp\left[-\boldsymbol{\sigma}^T \cdot \boldsymbol{b}^{(v)}\right] \prod_{j=1,\ldots,N} 2\cosh\left(b_j^{(h)} + \boldsymbol{W}^{(j)}\boldsymbol{\sigma}\right). \tag{5.9}$$

Note that the ansatz parameters are $\boldsymbol{v} = (\boldsymbol{b}^{(v)}, \boldsymbol{b}^{(h)}, W)$, with $\boldsymbol{b}^{(v)} \in \mathbb{C}^N$, $\boldsymbol{b}^{(h)} \in \mathbb{C}^M$ and $W \in \mathbb{C}^{N\times M}$. Complex weights have been considered, as opposed to the real weights of a standard RBM, because the wavefunction is complex-valued. We remark that the theorems introduced in Section 5.2.2 stating that in the limit of $M \to \infty$ this network can approximate any distribution, are not valid for the generic complex-valued case. However it is conjectured that they should hold, and it has been shown for some specific complex functions [244].

An alternative and more rigorous approach involves splitting the wavefunction into its modulus, and the phase, and represent each of those two different functionals with an

[4] It should be possible to further generalise the principle to continuous (infinite) Hilbert spaces, but the numerical implementation of the algorithm would have to undergo some non-trivial changes that, to the best of our knowledge, nobody has addressed yet.

identical neural network, namely

$$\psi_{\boldsymbol{v}}(\boldsymbol{\sigma}) = \sqrt{\mathcal{P}_{\boldsymbol{v}_A}(\boldsymbol{\sigma})}\exp[-1/2\log(\mathcal{P}_{\boldsymbol{v}_\theta}(\boldsymbol{\sigma}))], \tag{5.10}$$

where the amplitude $\mathcal{P}_{\boldsymbol{v}_A}$ and phase $\mathcal{P}_{\boldsymbol{v}_\theta}$ of the wavefunction are represented by two Boltzmann machines. In this case the two functionals are real-valued, therefore the representability theorems hold. Both those representations have been used in the literature, and it is still an open question whether one performs better than the other.

A limit of energy-based networks is that, even if they output valid wavefunctions, those wavefunctions are unnormalised and one typically has to compute the total amplitude to normalise observables, according to the definition

$$Z_{\boldsymbol{v}} = \sum_{\sigma\in\mathcal{H}} \psi_{\boldsymbol{v}}(\boldsymbol{\sigma}). \tag{5.11}$$

As the Hilbert space dimension can be huge, computing Z has exponential complexity and cannot be done in general. Therefore it is typically approximated through Markov-Chain sampling schemes which we will discuss later in the chapter [242].

Other ansätze

Even though the ansatz presented above has demonstrated a remarkable success, in recent months many more neural-network based ansätze have been proposed. In particular, *deep* FFNNs schemes have been shown to better handle bosonic systems [245, 246], and convolutional networks inspired by Google's PixelNet [247] have been proposed for directly sampling a normalised wavefunction [248]. As the focus of this manuscript is open quantum systems and not closed systems, we will not discuss them.

5.3.2 Density Matrix ansatz

To deal with open quantum systems, we do not need to parametrize the wavefunction but the full density matrix $\hat{\rho} \in \mathcal{B}(\mathcal{H})$

$$\hat{\rho} = \sum_{\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}}} \rho(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}}) \left|\boldsymbol{\sigma}\right\rangle \left\langle\tilde{\boldsymbol{\sigma}}\right|. \tag{5.12}$$

Taking inspiration from the Choi's isomorphism, one can easily perform the indentification $\mathcal{B}(\mathcal{H}) \leftrightarrow \mathcal{H}\otimes\mathcal{H}$. This is the same procedure used in the Matrix Product Operator representation of operators, and relies on representing them as vectors in a quadratically-bigger vector space. While this approach is viable, it encodes all linear operators (therefore all matrices) on $\mathcal{H}$, and is not restricted to density operators, which are positive-semidefinite. To construct a network that only parametrizes valid (positive-semidefinite) states, we resort to a construction based on the purification ansatz, originally proposed by Torlai et Melko [249].

In order to construct our neural network ansatz for the density matrix, we consider an extended Hilbert space $\mathcal{H} = \mathcal{H}_S \otimes \mathcal{H}_A$ where $\mathcal{H}_{S,A}$ represents respectively the system and ancillary Hilbert spaces. Such extended space is spanned by the basis set $\{|\boldsymbol{\sigma},\boldsymbol{a}\rangle\}$ where

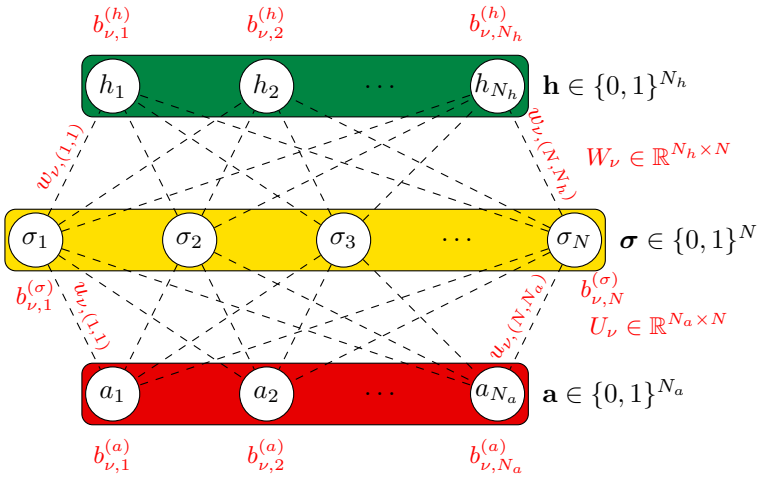

Figure 5.4: Graph representation of the purified state ansatz as described by the the energy functional in Eq. (5.16). Even if it is depicted differently, this ansatz is equivalent to a Restricted Bolzman Machine, where the latent space is made of the hidden and ancilla layers.

$\boldsymbol{a} = (a_1, a_2, ..., a_{N_a})$ labels the ancillary degrees of freedom. We start by considering a pure state in the extended Hilbert space, represented by the wavefunction $\psi_{\boldsymbol{v}}(\boldsymbol{\sigma}, \boldsymbol{a})$. In this framework the reduced density matrix of the system S is obtained by tracing out the ancillary degrees freedom, namely

$$\rho_{\boldsymbol{v}}(\boldsymbol{\sigma}, \boldsymbol{\sigma}') = \sum_{\boldsymbol{a}} \psi_{\boldsymbol{v}}(\boldsymbol{\sigma}, \boldsymbol{a}) \psi^{\star}_{\boldsymbol{v}}(\boldsymbol{\sigma}', \boldsymbol{a}). \tag{5.13}$$

This purified procedure automatically ensures that $\hat{\rho}_{\boldsymbol{v}}$ is Hermitian and positive semi-definite, as required for a density matrix.

The next step is to represent $\psi_{\boldsymbol{v}}(\boldsymbol{\sigma}, \boldsymbol{a})$ via a neural network ansatz such as that described in Section 5.3.1. The only difference is that the ancilla variables are encoded as a latent space that is not immediately traced away as the hidden space. Following the approach of Torlai and Melko [249] the complex nature of the purified wavefunctions is encoded by using two different real valued networks, one encoding the log-amplitude and the other the phase, namely

$$\psi_{\boldsymbol{v}}(\boldsymbol{\sigma}, \boldsymbol{a}) = \sqrt{\mathcal{P}_{\boldsymbol{v}_A}(\boldsymbol{\sigma}, \boldsymbol{a})} \exp[-1/2 \log(\mathcal{P}_{\boldsymbol{v}_\theta}(\boldsymbol{\sigma}, \boldsymbol{a}))]. \tag{5.14}$$

With this approach, both the amplitude $\mathcal{P}_{\boldsymbol{v}_A}(\boldsymbol{\sigma}, \boldsymbol{a})$ and phase $\mathcal{P}_{\boldsymbol{v}_\theta}(\boldsymbol{\sigma}, \boldsymbol{a})$ of the purified wavefunction are given by the Boltzmann-like expression

$$\mathcal{P}_\nu(\boldsymbol{\sigma}, \boldsymbol{a}) = \sum_{\boldsymbol{h}} \exp[-E_\nu(\boldsymbol{\sigma}, \boldsymbol{a}, \boldsymbol{h})], \tag{5.15}$$

with $\nu \in \{\boldsymbol{v}_A, \boldsymbol{v}_\theta\}$ labeling the sets of variational parameters for the two RBMs. The associated dimensionless energy reads

$$E_\nu(\boldsymbol{\sigma}, \boldsymbol{a}, \boldsymbol{h}) = \boldsymbol{\sigma} \cdot \mathbf{b}_\nu^{(\sigma)} + \boldsymbol{a} \cdot \boldsymbol{b}_\nu^{(a)} + \boldsymbol{h} \cdot \boldsymbol{b}_\nu^{(h)} + \boldsymbol{\sigma}^T W_\nu \boldsymbol{h} + \boldsymbol{\sigma}^T U_\nu \boldsymbol{a}. \tag{5.16}$$

Note that the ansatz parameters are $\boldsymbol{v} = (\boldsymbol{v}_A, \boldsymbol{v}_\theta)$ where $\boldsymbol{v}_\nu = (\mathbf{b}_\nu^{(\sigma)}, \boldsymbol{b}_\nu^{(a)}, \boldsymbol{b}_\nu^{(h)}, W_\nu, U_\nu)$. The rectangular matrix W_ν weighs the connections between the system variables (visible layer) to the auxiliary variables (hidden layer), while the weight matrix U_ν quantifies the connection between the system variables and the ancillary ones (ancillary layer). Such neural network ansatz is represented by a tri-partite Restricted Boltzmann Machine depicted in Fig. 5.4. In other words, there are two independent artificial neural networks, one for the amplitude ($\nu = A$) and one for the phase ($\nu = \theta$). By substituting those formulas into Eq. (5.13) and carrying out the sum over the ancillary degrees of freedom one obtains a closed formula for the entries of the density matrix:

$$\rho_{\boldsymbol{v}}(\boldsymbol{\sigma}, \boldsymbol{\sigma}') = \exp\left[\Gamma_{\boldsymbol{v}}^-(\boldsymbol{\sigma}, \boldsymbol{\sigma}') + \Gamma_{\boldsymbol{v}}^+(\boldsymbol{\sigma}, \boldsymbol{\sigma}') + \Pi_{\boldsymbol{v}}(\boldsymbol{\sigma}, \boldsymbol{\sigma}')\right] \tag{5.17}$$

where the expression of $\Gamma^{+/-}$ and Π are given by

$$\Gamma_{\boldsymbol{v}}^+(\boldsymbol{\sigma}, \tilde{\boldsymbol{\sigma}}) = \frac{1}{2}\left[\boldsymbol{b}_{\boldsymbol{A}}^{(\boldsymbol{v})} \cdot (\boldsymbol{\sigma} + \tilde{\boldsymbol{\sigma}}) + \sum_j \log\left[\mathcal{G}(\theta_A^{[j]}(\boldsymbol{\sigma}))\right] + \sum_j \log\left[\mathcal{G}(\theta_A^{[j]}(\tilde{\boldsymbol{\sigma}}))\right]\right], \tag{5.18}$$

$$\Gamma_{\boldsymbol{v}}^-(\boldsymbol{\sigma}, \tilde{\boldsymbol{\sigma}}) = \frac{i}{2}\left\{\boldsymbol{b}_{\boldsymbol{\theta}}^{(\boldsymbol{v})} \cdot (\boldsymbol{\sigma} - \tilde{\boldsymbol{\sigma}}) + \sum_j \log\left[\mathcal{G}(\theta_\theta^{[j]}(\boldsymbol{\sigma}))\right] - \sum_j \log\left[\mathcal{G}(\theta_\theta^{[j]}(\tilde{\boldsymbol{\sigma}}))\right]\right\}, \tag{5.19}$$

$$\Pi_{\boldsymbol{v}}(\boldsymbol{\sigma}, \tilde{\boldsymbol{\sigma}}) = \sum_k \log\left\{\mathcal{G}\left[1/2(\boldsymbol{\sigma} + \tilde{\boldsymbol{\sigma}})^T \boldsymbol{U}_{\boldsymbol{A}}^{[\boldsymbol{k}]} + i/2(\boldsymbol{\sigma} - \tilde{\boldsymbol{\sigma}})^T \boldsymbol{U}_{\boldsymbol{\theta}}^{[\boldsymbol{k}]} + b_{A,k}^{(a)}\right]\right\}. \tag{5.20}$$

$U_\nu^{[k]}$ is the $k-$th column of the U_ν matrix, and $\theta_\nu^{[j]}$ is the $j-$th component of the θ_ν vector

$$\boldsymbol{\theta}_{\boldsymbol{\nu}}(\boldsymbol{\sigma}) = \boldsymbol{b}_{\boldsymbol{\nu}}^{(\boldsymbol{h})} + \boldsymbol{\sigma}^T W_\nu. \tag{5.21}$$

$\mathcal{G}(x)$ is a non-linear activation function depending on the values that hidden spins can take. We considered binary spins ($h_i = \{0, 1\}$), so $\mathcal{G}(x) = 1 + e^{-x}$, but one can take $\mathcal{G}(x)$ to be any valid activation function. The representation power [250, 251, 252] of this ansatz can be systematically improved by increasing the density of the hidden ($\alpha = N_h/N$) and ancillary layer ($\beta = N_a/N$). The total number of real variational parameters of the neural network ansatz (the length of the vector $\boldsymbol{v}$) is given by the following formula:

$$N_{\text{par}} = \dim(\boldsymbol{v}) = 2(N + \alpha N + \beta N + \alpha N^2 + \beta N^2), \tag{5.22}$$

For $N \gg 1$ and assuming $\beta = 1$ (the motivation for this choice is discussed in Section 5.6), we get $N_{\text{par}} \sim \mathcal{O}\left((\alpha + 1)N^2\right)$. The number of parameters can be compared to a Matrix Product Operator representation of the density matrix where $N_{\text{par}}^{\text{MPO}} \sim \mathcal{O}(N\chi^2)$ where χ is the bond-link dimension [65], playing a role similar to α. For comparison, an exact description in the full Hilbert space requires $\mathcal{O}(2^{2N})$ parameters

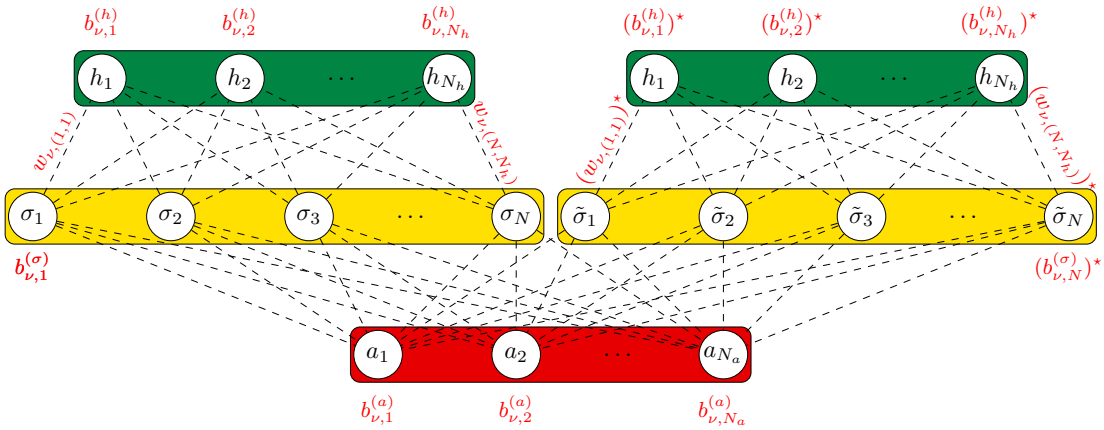

Figure 5.5: Graph representation of the neural density matrix, obtained by tracing over the ancilla, according to Eq. (5.13), of two coupled purified states as those in Fig. 5.4. The expression of this network is given by Eq. (5.17).

It is worth pointing out that the neural density matrix ansatz is not specific to this network topology, but relies only on the general fact that if two visible layers are connected by a shallow ancillary layer, the ancilla can be traced out analytically and an efficient neural-network description of the density matrix can be obtained. In fact, in a graph representation, the network, after tracing out has the structure shown in Fig. 5.5. The analytical trace can only be performed when the ancilla (red layer) is shallow. However the hidden space (green layer) does not need to be necessarily traced out analytically, so one can in general use a similar ansatz, with a deeper structure in place of the hidden layer. Mathematically, this means that the analytical trace necessary to perform the purification procedure allows to describe any state

$$\rho(\boldsymbol{\sigma}, \tilde{\boldsymbol{\sigma}}) = \exp[\Psi_0(\boldsymbol{\sigma}) + \Psi_0^\star(\tilde{\boldsymbol{\sigma}}) + \Pi_{\boldsymbol{v}}(\boldsymbol{\sigma}, \tilde{\boldsymbol{\sigma}})], \tag{5.23}$$

where $\Psi_0(\boldsymbol{\sigma})$ is an arbitrary deep neural network.

5.4 Variational method for open quantum systems

In quantum mechanics the variational method is a technique used to find approximately the ground state and low-lying excited states of a quantum Hamiltonian. The method uses an ansatz, which is a parametrised function depending on the set of parameters $\boldsymbol{v}$. The target state is determined by the optimal value of $\boldsymbol{v}_{gs}$, obtained by iteratively minimising the energy

$$E(\boldsymbol{v}) = \langle \psi_{\boldsymbol{v}} | \hat{H} | \psi_{\boldsymbol{v}} \rangle \geq \langle \psi_{gs} | \hat{H} | \psi_{gs} \rangle = E_{gs}, \tag{5.24}$$

which is a real quantity bounded from below by the ground-state energy. It is worth noting that $E(\boldsymbol{v})$ can be interpreted as a cost function to be minimised, in a similar spirit to how unsupervised learning minimises the cross-entropy with the target distribution.

Even if the ansatz we introduced in the previous section has real parameters $\mathcal{W}$, in general ansätze used in quantum mechanics will have complex weights, and the function to minimise will be

$$\mathcal{C} : \mathbb{C} \to \mathbb{R} \tag{5.25}$$
$$\boldsymbol{v} \to \mathcal{C}(\boldsymbol{v}), \tag{5.26}$$

which in general is a complex non-holomorphic function[5]. Unfortunately, the literature on complex functions is mainly focused on holomorphic functions, and the literature about variational methods in quantum mechanics deal mainly with real weights. In Section C.2 we introduce a few results on $\mathbb{CR}-$calculus, necessary to understand how to optimise non-holomorphic functions. If a reader is not interested in those details, the only result necessary to understand the following is that the steepest gradient of a real-valued complex function is the conjugate-gradient $\nabla_{\boldsymbol{v}^\star}$ and not the gradient $\nabla_{\boldsymbol{v}}$.

In this section we will present general formulas valid in the case of real-valued or holomorphic functions. In Section C.3 we will re-derive the same formulas, in the case of non-holomorphic ansätze.

5.4.1 Variational principles and cost functions

As we have shown in the previous section it is possible to write a variational ansatz for the density matrix by employing neural networks. We want to investigate the generalisation of the variational method to the case of open quantum systems, in order to find the optimal set of weights $\boldsymbol{v}_{ss}$ that best approximates the steady state of the ordinary differential equation

$$\dot{\hat{\rho}} = \mathcal{L}\hat{\rho}, \tag{5.27}$$

which is defined as $\hat{\rho}_{ss} = \lim_{t\to\infty} \hat{\rho}(t)$. We recall that it satisfies the condition $\mathcal{L}\hat{\rho}_{ss} = 0$.

A variational principle can be determined by a function $\mathcal{C} : \mathcal{W} \to \mathbb{R}$ with the following properties:

i) $\mathcal{C}(\boldsymbol{v}_{ss}) = c_{min} \iff \rho_{\boldsymbol{v}_{ss}} = \rho_{ss}$;

ii) $\forall \boldsymbol{v} \in \mathcal{W}, \quad \mathcal{C}(\boldsymbol{v}) \geq c_{min}$.

Moreover, in order to build an efficient method, it is also useful that:

iii) The computational cost of evaluating $\mathcal{C}(\boldsymbol{v})$ can be bounded through controlled approximations.

iv) The computational cost of evaluating the gradient of $\mathcal{C}(\boldsymbol{v})$ can be bounded through controlled approximations.

We will refer to $\mathcal{C}(\boldsymbol{v})$ as the cost-function, in order to highlight the similarity to the machine learning problem.

The distance in matrix between the time derivative of the state $\hat{\rho}(\boldsymbol{v}_{ss})$ and the time derivative of the state $\hat{\rho}(\boldsymbol{v})$ is a good candidate for the cost function, as it respects the first

[5] We recall that any real-valued function of complex parameters is non-holomorphic.

two points above. Because $\mathcal{L}\hat{\rho}(\boldsymbol{v}_{ss}) = 0$ the distance from the origin of $\mathcal{L}\hat{\rho}(\boldsymbol{v})$ corresponds with it's norm. Even if there exist several matrix norms, we will consider only Schatten p-norms because they are invariant under unitary transformations ($||\hat{\rho}||_i = ||U^\dagger \hat{\rho} U||_i$), preserving this feature of the density matrix [253]. For any value $p \in [1, \infty)$ the p-norm is defined as

$$\left\|\hat{A}\right\| = \left(\sum_i \lambda_i^p\right)^{1/p}, \tag{5.28}$$

where λ_i are the singular values of $\hat{A}$. The most common values are $p = 1$, called Trace-class norm, which is the sum of all singular values and $p = 2$ or the Hilbert–Schmidt norm, which is the sum of the squares of all entries. We define the cost function as

$$\mathcal{C}_{p,pn}(\boldsymbol{v}) = \frac{\|\mathcal{L}\hat{\rho}\|_p}{\|\hat{\rho}\|_{pn}}, \tag{5.29}$$

where p and pn select the Schatten norm used. By construction, $\forall p, pn$, Eq. (5.29) satisfies points (i) and (ii) above, therefore generating a good variational principle.

In his work, H. Weimer [125] analyses the case where $pn = 1$. This case is the most physically-relevant, as $pn = 1$ corresponds to normalising density matrices so that they have trace equal to 1. In the supplemental material of [125], he shows that the cost functions $\forall p \neq 1$ are biased, because they are exponentially small for the maximally mixed state.

If we consider an iterative procedure to minimise $\mathcal{C}_{1,1}(\boldsymbol{v})$, we need to compute its gradient. Assuming the decomposition $\hat{\rho}_{\boldsymbol{v}} = U_{\boldsymbol{v}} \Sigma_{\boldsymbol{v}} D_{\boldsymbol{v}}$ and $\mathcal{L}\hat{\rho}_{\boldsymbol{v}} = \tilde{U}_{\boldsymbol{v}} \tilde{\Sigma}_{\boldsymbol{v}} \tilde{D}_{\boldsymbol{v}}$,

$$\nabla_{\boldsymbol{v}} \mathcal{C}_{1,1}(\boldsymbol{v}) \propto \frac{(\tilde{U}_{\boldsymbol{v}} \tilde{D}_{\boldsymbol{v}}^\dagger)^\dagger \mathcal{L}\left[\nabla_{\boldsymbol{v}} \rho_{\boldsymbol{v}}\right]}{\mathrm{Tr}\left[\tilde{\Sigma}_{\boldsymbol{v}}\right]}. \tag{5.30}$$

Unfortunately, computing the matrices $\tilde{U}$ and $\tilde{V}$ requires performing the singular-value decomposition of $\mathcal{L}\rho_{\boldsymbol{v}}$, which is a computationally demanding task. The computational complexity of the SVD decomposition for a $l^N \times l^N$ matrix is $\mathcal{O}(l^{3N})$. It is possible to lower this cost by not computing the exact decomposition of $\mathcal{L}\hat{\rho}_{\boldsymbol{v}}$, but instead by using a low-rank approximation. The low-rank approximately sampled SVD decomposition is a very recent development in the field of numerical linear algebra [254, 255, 256], originally developed to solve the so-called Netflix problem. Unfortunately, the rank of a density matrix gives an upper bound on the Von-Neumann Entropy of the system, therefore performing a low-rank approximation effectively bounds the maximum entropy. Investigating the combination of variational methods with approximate SVD decomposition would be a very interesting research direction, but a method to treat states with limited entropy has already been developed in our group [69], and we desired to develop a technique that could complement the corner space renormalisation method. As such, we have decided not to consider the trace norm for the cost function.

Another norm, much easier to work with, is the L2 norm. The cost function

$$\mathcal{C}_{2,2}(\boldsymbol{v}) = \frac{\|\mathcal{L}\hat{\rho}_{\boldsymbol{v}}\|}{\|\hat{\rho}_{\boldsymbol{v}}\|} = \frac{\mathrm{Tr}\left[\hat{\rho}_{\boldsymbol{v}}^{\dagger}\mathcal{L}^{\dagger}\mathcal{L}\hat{\rho}_{\boldsymbol{v}}\right]}{\mathrm{Tr}\left[\hat{\rho}_{\boldsymbol{v}}^{\dagger}\hat{\rho}_{\boldsymbol{v}}\right]}, \tag{5.31}$$

does not suffer from the bias towards the maximally mixed state because it is normalised with the same L2 norm. It is easy to see that this cost function can be rewritten, with a bit of algebra, into the form

$$\mathcal{C}_{2,2}(\boldsymbol{v}) = \sum_{\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}}} p_{\boldsymbol{v}}(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}})\left|\mathcal{C}^{\mathrm{loc}}(\boldsymbol{v},\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}})\right|^2, \tag{5.32}$$

where:

$$\mathcal{C}^{\mathrm{loc}}(\boldsymbol{v},\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}}) = \sum_{\boldsymbol{\sigma}',\tilde{\boldsymbol{\sigma}}'} \mathcal{L}(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}},\boldsymbol{\sigma}',\tilde{\boldsymbol{\sigma}}')\frac{\rho_{\boldsymbol{v}}(\boldsymbol{\sigma}',\tilde{\boldsymbol{\sigma}}')}{\rho_{\boldsymbol{v}}(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}})}, \tag{5.33}$$

$$p_{\boldsymbol{v}}(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}}) = \frac{|\rho_{\boldsymbol{v}}(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}})|^2}{Z}, \tag{5.34}$$

$$Z = \sum_{\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}}} |\rho_{\boldsymbol{v}}(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}})|^2. \tag{5.35}$$

In particular, $\mathcal{C}^{\mathrm{loc}}(\boldsymbol{v},\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}})$ is the contribution to the cost function of the $(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}})$ entry, $p_{\boldsymbol{v}}(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}})$ is a well-defined probability and Z is the normalisation factor. Equation (5.32) shows explicitly that the cost function can be rewritten as the expectation value of the function $\mathcal{C}^{\mathrm{loc}}$ over the random variable $(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}})$. We remark that if $\mathcal{L}$ is K-Local, then only a polynomial number of entries for each row will be non-zero and computing $\mathcal{C}^{\mathrm{loc}}(\boldsymbol{v},\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}})$ will have a bounded cost. Any physical system with finite-range interaction and dissipation will have a K-local Liouvillian, therefore this is not a strong assumption.

In the following we will use notation $\left\langle \hat{A} \right\rangle_{\boldsymbol{v}}$ to indicate the expectation value of $A(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}})$, assumed to be a random variable on the space $\mathcal{H}\otimes\mathcal{H}$, weighted by the probability distribution $p_{\boldsymbol{v}}(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}})$. With that notation, we can express the cost function as the expectation value of the square modulus of the local cost function,

$$\mathcal{C}_{2,2}(\boldsymbol{v}) = \langle|\mathcal{C}^{\mathrm{loc}}_{\boldsymbol{v}}|^2\rangle_{\boldsymbol{v}}\,, \tag{5.36}$$

which this follows directly from Eq. (5.32).

An important quantity to compute when performing the stochastic optimisation of the variational ansätze is the gradient of the cost function. To provide more compact expressions, we express the diagonal log-derivative and conjugate-log-derivative operator in terms of its entries[6],

$$\mathcal{O}_{\boldsymbol{v}}(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}}) = \frac{\nabla_{\boldsymbol{v}}\hat{\rho}_{\boldsymbol{v}}(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}})}{\hat{\rho}_{\boldsymbol{v}}(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}})}, \qquad \mathcal{O}_{\boldsymbol{v}^\star}(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}}) = \frac{\nabla_{\boldsymbol{v}^\star}\hat{\rho}_{\boldsymbol{v}}(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}})}{\hat{\rho}_{\boldsymbol{v}}(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}})}, \tag{5.37}$$

[6]Technically, if we want to maintain the analogous to the variational method in closed Hamiltonian systems, $\mathcal{O}(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}},\boldsymbol{\sigma}',\tilde{\boldsymbol{\sigma}}')$ is a diagonal superoperator. As the only non-zero entries of $\mathcal{O}$ are those where the first two indices equal the last two, we can treat it as an operator.

where $\nabla_{\boldsymbol{v}}\hat{\rho}_{\boldsymbol{v}}(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}})$ is the $n_{\text{par}}-$dimensional gradient of the ansatz ($n_{\text{par}} = \dim[\mathcal{W}]$ being the number of variational parameters). As with the local cost function, this can also be understood as a random variable on the space $\mathcal{H}\otimes\mathcal{H}$. With the aid of the log-derivative we can express the gradient of the cost function as (details of the calculations are reported in Section C.3),

$$\nabla_{\boldsymbol{v}^\star}\mathcal{C}_{2,2}(\boldsymbol{v}) = \langle \mathcal{C}^{\text{loc}}\nabla_{\boldsymbol{v}^\star}\mathcal{C}^{\text{loc}}\rangle_{\boldsymbol{v}} - \langle|\mathcal{C}^{\text{loc}}|^2\rangle_{\boldsymbol{v}}\,\langle\mathcal{O}^\star_{\boldsymbol{v}}\rangle_{\boldsymbol{v}}\,, \tag{5.38}$$

where we assumed that $\hat{\rho}_{\boldsymbol{v}}(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}})$ is holomorphic with complex weights. If the ansatz is real-analytic and all weights are real, it is easy to show that the expression becomes

$$\nabla_{\boldsymbol{v}^\star}\mathcal{C}_{2,2}(\boldsymbol{v}) = \mathbb{R}e\left[\langle \mathcal{C}^{\text{loc}}\nabla_{\boldsymbol{v}^\star}\mathcal{C}^{\text{loc}}\rangle_{\boldsymbol{v}} - \langle|\mathcal{C}^{\text{loc}}|^2\rangle_{\boldsymbol{v}}\,\langle\mathcal{O}^\star_{\boldsymbol{v}}\rangle_{\boldsymbol{v}}\right]. \tag{5.39}$$

A more general form for the case of non-holomorphic ansätze can be found in the appendix. In both expressions above the gradient of the local cost function is defined as,

$$\nabla_{\boldsymbol{v}^\star}\mathcal{C}^{\text{loc}}(\boldsymbol{v},\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}}) = \sum_{\boldsymbol{\sigma}',\tilde{\boldsymbol{\sigma}}'}\mathcal{L}(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}};\boldsymbol{\sigma}',\tilde{\boldsymbol{\sigma}}')\frac{\rho_{\boldsymbol{v}}(\boldsymbol{\sigma}',\tilde{\boldsymbol{\sigma}}')}{\rho_{\boldsymbol{v}}(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}})}\mathcal{O}_{\boldsymbol{v}}(\boldsymbol{\sigma}',\tilde{\boldsymbol{\sigma}}'). \tag{5.40}$$

5.4.2 Sampling the cost function

The cost function of Eq. (5.32) is obtained by summing over the whole space $\mathcal{H}\otimes\mathcal{H}$, which is a task with exponential complexity. Among all local contributions to the cost-function, many states $(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}})$ have a negligible probability $p_{\boldsymbol{v}}(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}})\approx 0$ and could be dropped from the summation. If the probability distribution $p_{\boldsymbol{v}}(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}})$ is sufficiently peaked around a finite support, it is possible to compute the expectation values for the cost function and its gradient (Eqs. (5.36) and (5.39)) by stochastically sampling only a bounded number of elements according to the distribution.

Even though treating the cost function as a stochastic function bounds its computational cost, one must properly treat the sampling procedure itself. Sampling a probability distribution such as $p_{\boldsymbol{v}}(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}})$ requires in principle the computation of the full distribution, which involves the evaluation of an exponential number of elements. To avoid this, one might still randomly generate M elements $(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}})$ until he captures the majority of the distribution, so that $\sum_{i=1}^{M} p_{\boldsymbol{v}}(\boldsymbol{\sigma}_i,\tilde{\boldsymbol{\sigma}}_i)\approx Z$. For the particular ansatz we presented in the previous section, however, computing the normalisation factor Z requires performing the full sum over the doubled Hilbert space, an exponentially complex task. In general, it is not tractable to compute this quantity, except for a small class of autoregressive ansätze (built on convolutional neural networks originally proposed by Google) [248].

To avoid this problem, it is possible to use more advanced sampling techniques such as Markov Chains, which in the limit of sufficiently long chains provide good sampling of the space [257, 258].

5.4.3 Encoding symmetries

Physical systems often exhibit symmetries, which must also be present in their steady-state [38]. Symmetries can be exploited to reduce the computational complexity of computational methods or to improve their stability. Assume that the Lindblad master equation

is invariant under the action of a symmetry group defined by the set of linear transformations $\mathcal{T}_s$ where $s = 1, \dots S$ (S is the dimension of the symmetry). The density matrix $\hat{\rho}$ transforms according to

$$\hat{\rho} \to \mathcal{T}_s \hat{\rho} \mathcal{T}_s^\dagger, \tag{5.41}$$

and should be invariant under this transformation for every value of s.

It is possible to generalise certain neural network ansätze to be invariant under the action of translations or rotations [259, 260], as it has already been demonstrated in the case of closed Hamiltonian systems [70, 71]. This approach should be easy to generalise to the case of mixed states by ensuring that the same transformation is enforced both in the rows and the columns of the density matrix. While this approach is of interest because it reduces the number of parameters, it requires that the first layer be a dense layer and therefore is not completely general[7]. Moreover, it is not evident how to enforce more complicated symmetries such as $\mathbb{Z}_N$, $U(1)$ or other non-abelian symmetries.

A simpler yet more general approach that could be applied would be to symmetrize the evaluation of the network, by replacing every evaluation of $\rho(\boldsymbol{\sigma}, \tilde{\boldsymbol{\sigma}})$ with

$$\rho_{sym}(\boldsymbol{\sigma}, \tilde{\boldsymbol{\sigma}}) = \sum_{s=1}^{S} \rho(\mathcal{T}_s \boldsymbol{\sigma}, \mathcal{T}_s \tilde{\boldsymbol{\sigma}}), \tag{5.42}$$

and performing a similar transformation on the gradient $\mathcal{O}(\boldsymbol{\sigma}, \tilde{\boldsymbol{\sigma}})$. This technique has been applied in the context of convolutional ansätze of closed Hamiltonian systems [248]. The technique can also be improved by performing the Markov chain in the space of the irreducible representations, as recently reported by Vieijra et al. [261].

5.4.4 A note on alternative expressions for the cost function

We remark that the cost function $\mathcal{C}_{2,2}$ can be rewritten in another, possibly more intuitive form, namely

$$\mathcal{C}_{2,2}(\boldsymbol{v}) = \sum_{\boldsymbol{\sigma}, \tilde{\boldsymbol{\sigma}}} p_{\boldsymbol{v}}(\boldsymbol{\sigma}, \tilde{\boldsymbol{\sigma}}) \mathcal{C}^{\text{loc, }2}(\boldsymbol{v}, \boldsymbol{\sigma}, \tilde{\boldsymbol{\sigma}}), \tag{5.43}$$

$$\mathcal{C}^{\text{loc, }2}(\boldsymbol{v}, \boldsymbol{\sigma}, \tilde{\boldsymbol{\sigma}}) = \sum_{\boldsymbol{\sigma}', \tilde{\boldsymbol{\sigma}}'} (\mathcal{L}^\dagger \mathcal{L})(\boldsymbol{\sigma}, \tilde{\boldsymbol{\sigma}}, \boldsymbol{\sigma}', \tilde{\boldsymbol{\sigma}}') \frac{\rho_{\boldsymbol{v}}(\boldsymbol{\sigma}', \tilde{\boldsymbol{\sigma}}')}{\rho_{\boldsymbol{v}}(\boldsymbol{\sigma}, \tilde{\boldsymbol{\sigma}})}. \tag{5.44}$$

The equivalence between the two expressions is valid only as long as the sum $\sum_{\boldsymbol{\sigma}, \tilde{\boldsymbol{\sigma}}}$ is performed exactly over the whole space. However, when the sum is evaluated via a Monte Carlo procedure, Eq. (5.32) and Eq. (5.43) have different convergence properties. In particular, Eq. (5.43) can assume a positive or negative real part. Equation (5.32), instead, is a sum over positive real terms, and respects the 0-variance property for the steady-state.

[7] Restricted Boltzmann machines, when written as a feedforward network have a single dense layer. The symmetric counterpart of a Boltzmann machine with αN hidden neurons is a single convolutional layer with α kernels. This reduces the number of parameters by a factor $S = N$ (Assuming all translation symmetries of a periodic lattice).

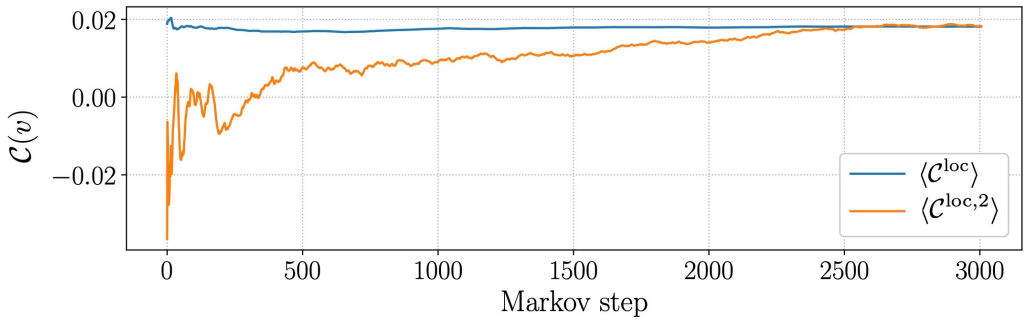

Figure 5.6: The cost function $\mathcal{C}_{2,2}$ computed iteratively with the two schemes of Eq. (5.32) and Eq. (5.43) respectively. While both schemes converge to the same value, the first one converges much faster to the exact value.

To make a comparison to closed systems, optimising the first is similar in spirit to performing a variance minimisation of the variational wavefunction as opposed to the latter, which is a simple energy minimisation [262]. This leads to faster convergence as a function of the Markov chain length when $\mathcal{C}_{2,2}$, as shown in Fig. 5.6. We remark that computing the gradient of $\mathcal{C}_{2,2}$ through this expression of local weights leads to a different form of the gradient,

$$\nabla_{\boldsymbol{v}^\star} \mathcal{C}_{2,2}(\boldsymbol{v}) = \langle \mathcal{C}^{\mathrm{loc},2}\ \mathcal{O}^\star_{\boldsymbol{v}} \rangle_{\boldsymbol{v}} - \langle \mathcal{C}^{\mathrm{loc},2} \rangle_{\boldsymbol{v}}\ \langle \mathcal{O}^\star_{\boldsymbol{v}} \rangle_{\boldsymbol{v}}\,, \tag{5.45}$$

which is obtained by imposing the fact that $\mathcal{L}^\dagger \mathcal{L}$ is hermitian. Since this super-operator is quadratically denser than $\mathcal{L}$ itself, it is usually more expensive to compute.

For those reasons, we strongly recommend not to use this expression of the local cost-function for driven-dissipative systems.

5.5 Optimisation algorithms for the variational problem

Finding the global minima $\boldsymbol{v}_{ss}$ of $\mathcal{C}(\boldsymbol{v})$ is an optimisation problem. To solve optimisation problems there exist mainly two classes of classical algorithms: gradient-free global optimisers [263], such as swarm algorithms [264] or simulated annealing [265], and gradient-based iterative methods. When optimising functions in very high dimensional spaces (dim $>$ 500) global optimisers tend to perform rather poorly, and the only possible choice are iterative methods.

5.5.1 Stochastic Gradient Descent

The simplest iterative method used when optimising (or *learning*) neural networks is the gradient descent method, a first-order iterative method that updates the weights according to the gradient at each iteration.

$$\boldsymbol{v}_{i+1} = \boldsymbol{v}_i - \eta \nabla_{\boldsymbol{v}^\star}\ \mathcal{C}(\boldsymbol{v}_i) \tag{5.46}$$

where η is the learning rate, and determines how fast one descends along the gradient direction. In the context of machine learning, it is often intractable to evaluate exactly the gradient $\nabla_{\boldsymbol{v}}\, \mathcal{C}(\boldsymbol{v})$, which is instead sampled stochastically; in that case the method is called *Stochastic* gradient descent (SGD) [266].

It is rather interesting to note that while exact gradient descent optimisation of a non-convex function might easily get stuck into local minima, in the typical problems of machine learning (stochastic optimisation and very high-dimensional spaces) this is not the case. In particular, it has been shown that local minima are rarely a problem in deep networks when the dimensionality is sufficiently high [267].

Thermalisation-like features

Stochastic gradient descent can, under certain assumptions[8], be interpreted as a Langevin process. Computing the gradient approximately introduces random noise that can be modelled as a Wiener process,

$$d\boldsymbol{v} = -\nabla_{\boldsymbol{v}^\star}\, \mathcal{C}(\boldsymbol{v})\eta + \sqrt{\beta^{-1}D(\boldsymbol{v})}dW, \tag{5.47}$$

where $\beta(\eta)$ is an effective temperature that depends on the sampling algorithm and $D(\boldsymbol{v})$ is the variance of $\nabla_{\boldsymbol{v}^\star}\, \mathcal{C}(\boldsymbol{v})$ [268]. Stochastic sampling of the gradient introduces an effective temperature that allows the optimisation to get out of local minima, in the same spirit of annealing. Evidently, a tradeoff is required between the accuracy of the solution, which increases with η and the speed with which one falls towards the minima, which depends on η^{-1}.

5.5.2 Nesterov accellerated momentum

An improvement to the basic stochastic gradient descent, is obtained by changing η dynamically. This is effective to accelerate the optimisation in regions where the gradient is very small. Moreover, it protects from sudden variations in the gradient, which sometime arise due to ill-conditioning of the problem or insufficient sampling.

A possible approach to obtain an effective adaptivity of η is by introducing momentum in the update rules, for example through the so-called Newton scheme. Here, we will detail a more advanced formulation known as *Nesterov* accelerated momentum, where the equation of motion have been integrated with the leap-frog scheme in order to guarantee a more stable convergence [269]. Within this approach the updates are given by the two iterative equations

$$\dot{\boldsymbol{v}}_{t+1} = \mu\boldsymbol{v}_t - \eta\nabla_{\boldsymbol{v}^\star}\, \mathcal{C}(\boldsymbol{v} + \mu\dot{\boldsymbol{v}}_t), \tag{5.48}$$

$$\boldsymbol{v}_{t+1} = \boldsymbol{v}_t + \dot{\boldsymbol{v}}_{t+1}, \tag{5.49}$$

where the initial speed can be set to $\dot{\boldsymbol{v}}_0 = 0$, $\mu \in [0, 1[$ is the damping coefficient and η is the bare learning rate. It is possible to show that the effective step size depends on the gradient and will be $\leq \eta\frac{\mu}{1-\mu}$.

[8] Mean field treatment of the different parameters, otherwise geometric noise terms might appear.

5.5.3 Stochastic reconfiguration

A different method that improves convergence towards the global minimum of the cost function is the stochastic reconfiguration method, also known as natural gradient descent by the machine learning community [270, 271]. The stochastic reconfiguration corrects the gradient descent so that small variations $\boldsymbol{\delta v}$ in the parameters do not cause large variations, according to a certain norm, for the density matrices. The generalisation of this method, previously developed for closed systems [243], can be straightforwardly made in our case. In what follows we give a sketch of the derivation.

Consider the first-order Taylor expansion of the variational density matrix with respect to the variational parameters:

$$\hat{\rho}_{\boldsymbol{v}+\boldsymbol{\delta v}} = \hat{\rho}_{\boldsymbol{v}} + \sum_i \delta v_i \mathcal{O}_{v_i} \hat{\rho}_{\boldsymbol{v}}, \tag{5.50}$$

where $i \in [1, N_{\text{par}}]$ spans all the variational parameters. The distance between $\hat{\rho}_{\boldsymbol{v}}$ and $\hat{\rho}_{\boldsymbol{v}+\boldsymbol{\delta v}}$ is defined to be:

$$\delta s^2_{\boldsymbol{v}} = \left\| \frac{\hat{\rho}_{\boldsymbol{v}+\boldsymbol{\delta v}}}{\|\hat{\rho}_{\boldsymbol{v}+\boldsymbol{\delta v}}\|} - \frac{\hat{\rho}_{\boldsymbol{v}}}{\|\hat{\rho}_{\boldsymbol{v}}\|} \right\|_2^2, \tag{5.51}$$

where in the following we have chosen the 2-norm, defined by $\|\hat{A}\|_2^2 = \mathrm{Tr}\left[\hat{A}^\dagger \hat{A}\right]$. It is also known that the distance between two vectors in a space is given by the metric tensor S as

$$\delta s^2_{\boldsymbol{v}} = \delta \boldsymbol{v}^\dagger S_{\boldsymbol{v}} \delta \boldsymbol{v}. \tag{5.52}$$

By keeping only the leading terms in $\delta \boldsymbol{v}$ it is possible to rewrite Eq. (5.51) in order to show that the metric tensor can be written as the expectation value (valid for an holomorphic ansatz)

$$S_{\boldsymbol{v}} = \langle \mathcal{O}_{\boldsymbol{v}} \mathcal{O}^\dagger_{\boldsymbol{v}} \rangle_{\boldsymbol{v}} - \langle \mathcal{O}_{\boldsymbol{v}} \rangle_{\boldsymbol{v}} \langle \mathcal{O}^\dagger_{\boldsymbol{v}} \rangle_{\boldsymbol{v}}. \tag{5.53}$$

In the case of an ansatz with real parameters this expression is instead

$$S_{\boldsymbol{v}} = \mathbb{R}e[\langle \mathcal{O}_{\boldsymbol{v}} \mathcal{O}^\dagger_{\boldsymbol{v}} \rangle_{\boldsymbol{v}}] - \mathbb{R}e[\langle \mathcal{O}_{\boldsymbol{v}} \rangle_{\boldsymbol{v}}] \mathbb{R}e[\langle \mathcal{O}^\dagger_{\boldsymbol{v}} \rangle_{\boldsymbol{v}}]. \tag{5.54}$$

For the more complicated case of non-holormophic ansätze check the appendix. We remark that it has the same form of the covariance matrix of the stochastic gradient, and therefore it is equivalent to the Fisher's (classical) information matrix for the distribution ρ.

We will now show how it is possible to use the information contained in the metric tensor to correct an otherwise standard gradient descent. Consider a generic iterative method; if the variational weights are updated according to the rule $\boldsymbol{v} \to \boldsymbol{v}' = \boldsymbol{v} + \delta \boldsymbol{v}$, after every iteration the change in the cost function will be

$$\Delta \mathcal{C} = \mathcal{C}(\boldsymbol{v} + \delta \boldsymbol{v}) - \mathcal{C}(\boldsymbol{v}) = \delta \boldsymbol{v}^T \nabla_{\boldsymbol{v}} \mathcal{C} + \delta \boldsymbol{v}^\dagger \nabla_{\boldsymbol{v}^\star} \mathcal{C} + \mathcal{O}(|\delta \boldsymbol{v}|^2), \tag{5.55}$$

where we only retained terms linear in the weight updates. To enforce the constrain that an update $\delta \boldsymbol{v}$ gives a small variation in the space of $\hat{\rho}$, one can introduce a Lagrange multiplier μ and minimise with respect to the variations of $\delta \boldsymbol{v}$ the quadratic form

$$\Delta \mathcal{C} + \mu \delta s^2 = \delta \boldsymbol{v}^T \nabla_{\boldsymbol{v}} \mathcal{C} + \delta \boldsymbol{v}^\dagger \nabla_{\boldsymbol{v}^\star} \mathcal{C} + \mathcal{O}(|\delta \boldsymbol{v}|^2) + \mu \delta \boldsymbol{v}^\dagger S_{\boldsymbol{v}} \delta \boldsymbol{v}, \tag{5.56}$$

where we have added to both sides of the equation the equality $\mu\delta s^2 = \mu\delta\boldsymbol{v}^\dagger S\delta\boldsymbol{v}$. By asking that the partial derivative with respect to $\delta\boldsymbol{v}^\dagger$ is zero, we obtain the solution

$$\nabla_{\boldsymbol{v}^\star}\mathcal{C} = -S_{\boldsymbol{v}}\delta\boldsymbol{v}. \tag{5.57}$$

By inverting S we then obtain the stochastic reconfiguration (or natural gradient) update rule,

$$\boldsymbol{v} \to \boldsymbol{v} + \delta\boldsymbol{v} \quad \text{where} \quad \delta\boldsymbol{v} = -S_{\boldsymbol{v}}^{-1}\nabla_{\boldsymbol{v}^\star}\mathcal{C}(\boldsymbol{v}), \tag{5.58}$$

where the metric tensor S is often called the SR matrix. Notably, this formula also describes the standard steepest-gradient descent, where the metric tensor is taken to be the identity, as is the case for the euclidean space $\mathcal{W}$. The metric tensor $S_{\boldsymbol{v}}$ is a real, dense, symmetric matrix of size $N_{\text{par}} \times N_{\text{par}}$, which has often many degenerate eigenvalues. Degenerate eigenvalues represent the fact that not all parameters are (linearly) independent around $\boldsymbol{v}$. Moreover, the null eigenvalues identify parameters, or combination of parameters, which do not change the resulting density matrix $\hat{\rho}$. Consequently, it is hard to numerically invert the metric tensor efficiently and accurately. Moreover, matrix inversion methods are computationally very expensive and their precision scales rather badly. When working with a neural network ansatz it is customary to have $N_{\text{par}} > 1000$. Krylov-space iterative solvers such as MINRES-QLP are more suitable at determining the update vector $\delta\boldsymbol{v}$ without ever inverting the matrix [272]. To avoid numerical stability issues arising from the zero eigenvalues of $S_{\boldsymbol{v}}$, we can add a uniform diagonal bias b,

$$(S_{\boldsymbol{v}} + \ b\ \mathbb{I})\delta\boldsymbol{v} = \nabla_{\boldsymbol{v}^\star}\mathcal{C}(\boldsymbol{v}), \tag{5.59}$$

with $b \in [10^{-4}, 10^{-2}]$ in practical numerical implementations. As this system is not equivalent to the original, the bias sets a lower bound to the lowest possible value of the cost function. As such, it should either be lowered during the optimisation, and ensure that it is always at least one order of magnitude smaller than the current $\mathcal{C}(\boldsymbol{v})$. An alternative approach employs a non-diagonal bias such as that used in Ref. [73].

We would also like to remark that, by using iterative methods one only needs to compute products between $S_{\boldsymbol{v}}$ and a vector used by the linear solver. As the matrix $S_{\boldsymbol{v}}$ is nether computed exactly but it is obtained by sampling the log-derivatives $\mathcal{O}_{\boldsymbol{v}}$, it is possible to avoid computing the full matrix altogether and implement the product via matrix-free methods. This would also lower the memory cost from N_{par}^2 to $N_{\text{par}} \times M$ where M is the number of stochastic samples. We also expect that the computational cost of a well-designed distributed implementation to be lower than the full-matrix equivalent, though we have not yet experimented with that.

5.6 Benchmarking on spin systems

In order to benchmark our neural-network approach for open quantum systems, we consider here the dissipative quantum transverse Ising model, whose Hamiltonian is

$$H = \frac{V}{4}\sum_{\langle j,l\rangle}\hat{\sigma}_j^z\hat{\sigma}_l^z + \frac{g}{2}\sum_j\hat{\sigma}_j^x, \tag{5.60}$$

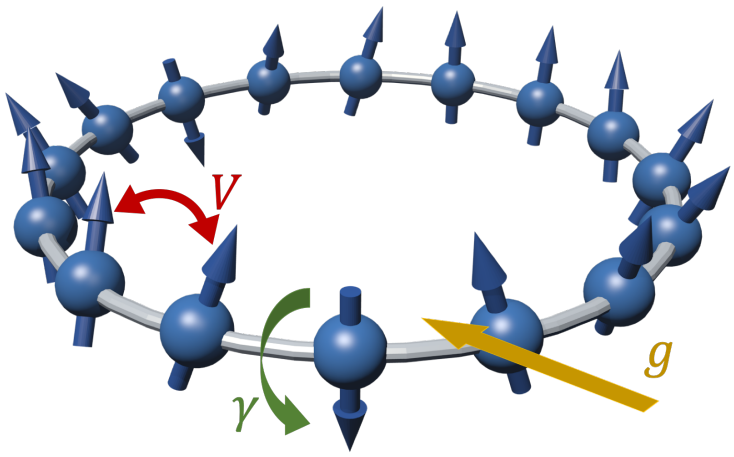

Figure 5.7: Top panel: sketch of the considered physical system described by the dissipative quantum transverse Ising 1D model with periodic boundary conditions. The quantity g denotes an applied magnetic field, V the spin-spin coupling and γ the spin flip rate.

being $\hat{\sigma}_j^\alpha$ the Pauli matrices ($\alpha \in \{x, y, z\}$) acting on the j-th site. The first term represents the nearest-neighbour spin-spin interaction depending only on the z-components, being V the coupling strength. The second term accounts for a local and uniform magnetic field along the transverse direction x. We consider local dissipative spin-flip processes described by the site-dependent jump operator $\hat{L}_j^{(z)} = \hat{\sigma}_j^- = \frac{1}{2}(\hat{\sigma}_j^x - i\hat{\sigma}_j^y)$, which fully determine the Lindblad master equation.

Numerical results for steady-state observables of the dissipative quantum transverse Ising model on a 1D periodic chain are reported in Fig. 5.8. An up to date package implementing the computational method is freely accessible in GitHub [273], while the precise version used for the simulations presented here is stored in Zenodo [274]. We discuss some implementation details are briefly discussed in Section C.4. In particular, we report the spatial components of the averaged magnetisation as a function of the magnetic field g (in units of the dissipation rate γ) for $V/\gamma = 2$. For $N = 16$ lattice sites the predictions of the neural-network variational method (circles) are compared to the results obtained with a brute-force exact integration of the master equation in the whole Hilbert space, showing a good agreement over all the parameter range. For $g \lesssim \gamma$ and $g \gtrsim 2.5\gamma$ a remarkable precision is reached for all the local observables with a low density of the hidden and ancillary layer $\alpha = \beta = 1$ and $\mathcal{O}(10^2)$ minimisation steps. For $1 \lesssim g/\gamma \lesssim 2.5$ an higher number of variational parameters is required. In particular, as shown in Fig. 5.9, a systematic improvement of the relative error $\epsilon_{\text{rel}}\big[\langle\sigma_x\rangle\big]$ with respect to the exact solution can be obtained by increasing the ancilla density β. This figure also suggests that varying the hidden layer density α has little effect on the representation power of the ansatz. Indeed, if one rewrites the ansatz in the form $\rho(\boldsymbol{\sigma}, \tilde{\boldsymbol{\sigma}}) = \exp[\Psi_0(\boldsymbol{\sigma}) + \Psi_0^\star(\tilde{\boldsymbol{\sigma}}) + \Pi_{\boldsymbol{v}}(\boldsymbol{\sigma}, \tilde{\boldsymbol{\sigma}})]$ it is easy to show that α increases the representation power of the pure-part Ψ_0 and that

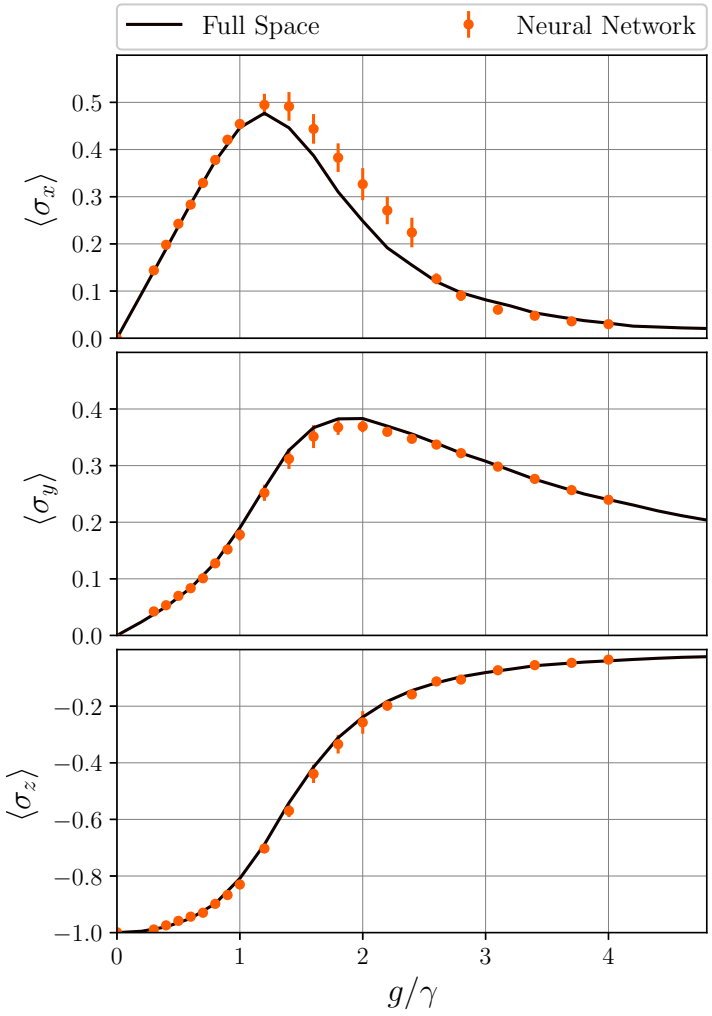

Figure 5.8: The prediction of the neural-network variational calculations (circles) are compared to the results obtained by quantum trajectory simulations of the master equation by considering the whole Hilbert space (solid lines). The top, middle and bottom panels depict the expectation values of the three components of the averaged magnetisation as a function of the applied magnetic field g (in units of γ). Model parameters: $V/\gamma = 2$ (spin-spin coupling), $N = 16$ (number of lattice sites). Neural-network parameters: $\alpha = \beta = 1$ for $g \leq \gamma$ and $g \geq 2.5\gamma$ while $\alpha = 1$ and $\beta = 4$ for the remaining points. The parameters required for the convergence of the Monte Carlo calculations depend on the value of g/γ, with the intermediate region being the most demanding. The maximum number of accepted Monte Carlo samples is 8640 and the maximum number of steps for the stochastic gradient descent is 10^4. For points outside the intermediate region 3000 accepted Monte Carlo samples and 10^3 iteration steps have been performed.

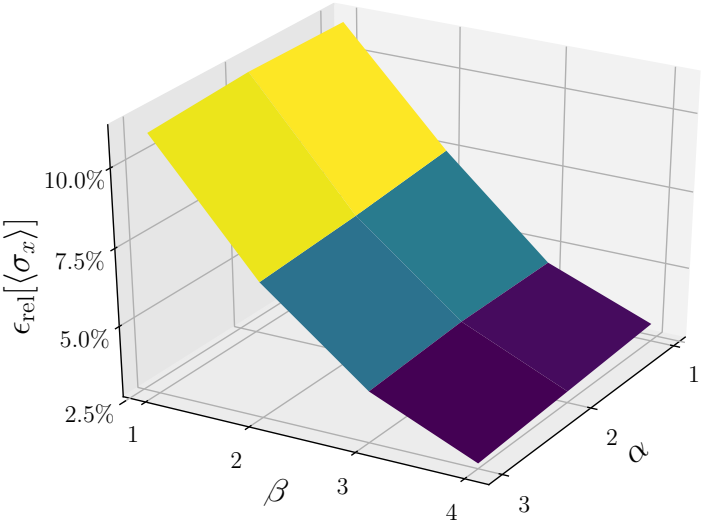

Figure 5.9: Relative error with respect to the exact result for the observable $\langle\sigma_x\rangle$ as a function of α and β. Parameters are set as in Fig.5.8 but for a fixed value $g/\gamma = 1.2$.

β increases the representation power of the mixed part II. This hints that the system is encoding all the information into the mixed part, which is unfortunately the part that is constrained due to the analytical integration of the ancilla and that cannot be made deeper.

Interestingly, for $1 \lesssim g/\gamma \lesssim 2.5$, we note that the gradient-descent procedure requires more iterations. This region corresponds to the range of g/γ where the smallest nonzero eigenvalue of $\mathcal{L}^\dagger\mathcal{L}$ decreases significantly [275]. In this range the steady-state density matrix also displays nontrivial correlations and non-thermal mixedness properties [275]. Remarkably, the fidelity of the reconstructed local density matrix with respect to the exact one is alway larger then 0.998 for all the values of g/γ considered.

Finally, as an example of convergence, the top-left panel of Fig. 5.10 depicts a typical evolution of the cost function in the iterative minimisation procedure for a fixed set of parameters ($g/\gamma = 1$), showing a good convergence towards the global minimum. In the bottom-left panel of Fig. 5.10, the convergence of the x-component of the averaged magnetisation is also reported. For this particular value of the parameters, good agreement in the local observables is attained when $\mathcal{C} \lesssim 10^{-2}$. It is however not easy to generalise this quantity to a general system, as this cost function does not scale homogeneously with the system size.

The convergence towards the steady state can also be represented more graphically thanks to mean field techniques. By taking the weights of the network at the end of the optimisation procedure $\boldsymbol{v}^\star$ and picking two gaussian random vectors $\boldsymbol{\delta}_\alpha$ and $\boldsymbol{\delta}_\beta$, we can compute the cost function $\mathcal{C}(\boldsymbol{v} + \alpha\boldsymbol{\delta}_\alpha + \beta\boldsymbol{\delta}_\beta)$. If $\boldsymbol{v}^\star$ is a local minima the landscape in the manifold spanned by the two random vectors will be representative [276], as shown in the

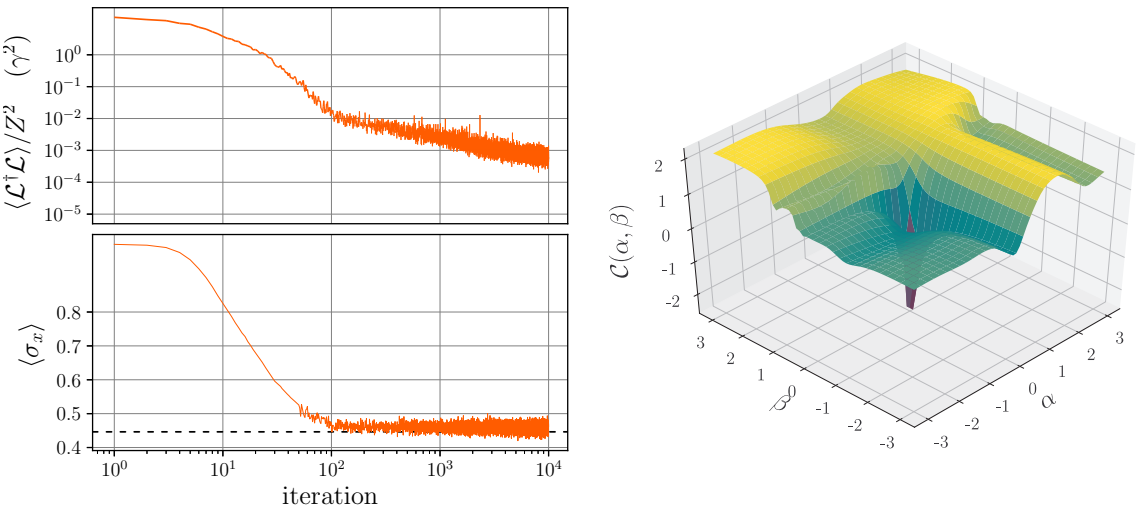

Figure 5.10: Left panel: same parameters as in Fig. 5.8 but for a fixed value $g/\gamma = 1$. Top panel: the cost function is shown as a function of the iteration steps. Bottom panel: the corresponding evolution of the x-component of the average magnetisation during the stochastic minimisation is shown. Right panel: landscape of the log-cost function $log[\mathcal{C}(\boldsymbol{v})]$ along two non-orthogonal representative directions α and β around the global minima $(\alpha, \beta) = 0$, computed according to [276].

right panel of Fig. 5.10.

To conclude, in Fig. 5.11 we report the scaling of the runtime performance of the neural network algorithm, compared to traditional brute-force Runge-Kutta integration of the master equation and to a quantum trajectory approach for the value of the transverse field $g/\gamma = 0.7$ and a number of monte-carlo samples depending loosely on the size of the system. We see an advantage to using Quantum trajectories for lattices with as little as 12 sites.

5.7 Possible future generalisation based on purified space sampling

To conclude the chapter we will introduce a possible improvement that could be implemented in the future to make our method even more general.

We have shown in the benchmarking that the network seems to learn even relatively-pure states through the components encoding the correlations among row and columns of the density matrix. If the neural density matrix ansatz is written in the form,

$$\rho(\boldsymbol{\sigma}, \tilde{\boldsymbol{\sigma}}) = \exp[\Psi_0(\boldsymbol{\sigma}) + \Psi_0^\star(\tilde{\boldsymbol{\sigma}}) + \Pi_{\boldsymbol{v}}(\boldsymbol{\sigma}, \tilde{\boldsymbol{\sigma}})], \tag{5.61}$$

the network is encoding information in $\Pi_{\boldsymbol{v}}$, which is the only part that we cannot change freely. If we want to write a ansatz for the density matrix $\rho(\boldsymbol{\sigma}, \tilde{\boldsymbol{\sigma}})$ starting from some

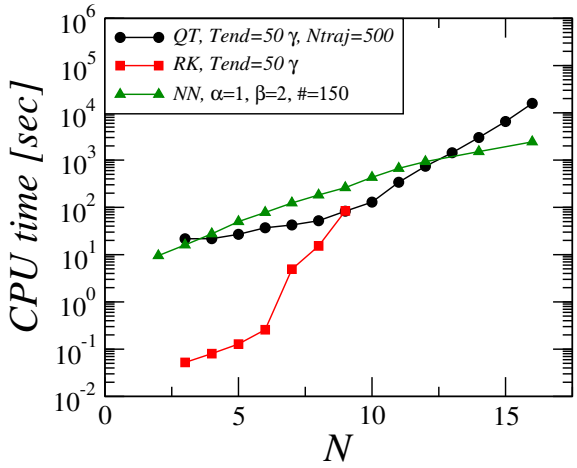

Figure 5.11: Runtime cost (in CPU-core-seconds) of various algorithms as a function of the system size N. Red squared: the brute-force RK45 integration of the master equation up to $T_{\text{end}} = 50\gamma$. Green triangles: the integration of 500 trajectories up to $T_{\text{end}} = 50\gamma$ with the TSIT5 ode-integrator. Black circles: the CPU-time of the Neural Network algorithm presented in this paper, with $\alpha = 1$, $\beta = 2$, $M = 10^3 * (1 + \log_2(N))$ Markov-chain samples and 150 iterations.

purified wavefunction $\psi(\boldsymbol{\sigma}, \boldsymbol{a})$, we need to perform the analytical summation over the ancillary degrees of freedom. We had initially thought this would not be a big limitation, as we could substitute the relatively simple form for Ψ_0 that we have used in the benchmark with more complicated functions, mainly inspired from Machine Learning literature. However, our results seem to suggest that this would not be efficient, as in the end the network would concentrate the encoding of the information into the part that is fixed.

In the following we propose a sampling and optimisation scheme that allows for truly generic and unconstrained ansätze, while still enforcing the positivity of the resulting density matrix. The idea stems from the purification ansatz, which is positive definite for an arbitrary wavefunction $\psi^\star_{\boldsymbol{v}}(\boldsymbol{\sigma}, \boldsymbol{a})$, we could compute the cost function by sampling also in the ancillary space, thus avoiding the analytical summation over the ancilla's variables.

To be able to sample stochastically the cost function by performing a Markov-Chain in the space $(\boldsymbol{\sigma}, \boldsymbol{a})$, where every configuration is weighted by $|\psi^\star_{\boldsymbol{v}}(\boldsymbol{\sigma}, \boldsymbol{a})|^2$, we must impose the proper normalisation. Interestingly, purified wavefunctions have the property that

$$\|\psi_{\boldsymbol{v}}\|^2 = \mathrm{Tr}[\hat{\rho}] = \sum_{\boldsymbol{\sigma},\boldsymbol{a}} |\psi^\star_{\boldsymbol{v}}(\boldsymbol{\sigma}, \boldsymbol{a})|^2. \tag{5.62}$$

Therefore the following cost function can be sampled in that space

$$\mathcal{C}_{2,1}(\boldsymbol{v}) = \frac{\|\mathcal{L}\hat{\rho}_{\boldsymbol{v}}\|_2}{\mathrm{Tr}[\hat{\rho}_{\boldsymbol{v}}]^2}. \tag{5.63}$$

We remark that this cost function, when computed for sufficiently big systems, would be biased towards the maximally mixed state because it is normalised with the same L2 norm [125]. Nevertheless, it is possible that this bias would not arise in iterative schemes

if properly initialised to stay far from the mixed state. This suggests that it might a viable alternative.

By substituting Eq. (5.62) into Eq. (5.63) we obtain

$$\mathcal{C}_{2,\mathrm{Tr}}(\boldsymbol{v}) = \sum_{\boldsymbol{\sigma_a},\boldsymbol{\sigma'_a}} p(\boldsymbol{\sigma_a})p(\boldsymbol{\sigma'_a}) \sum \left(\mathcal{L}(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}},\boldsymbol{\sigma'},\tilde{\boldsymbol{\sigma}'})\frac{\psi(\tilde{\boldsymbol{\sigma}}_{\boldsymbol{a}})}{\psi(\boldsymbol{\sigma_a})}\right)\left(\mathcal{L}(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}},\boldsymbol{\sigma'},\tilde{\boldsymbol{\sigma}'})\frac{\psi(\tilde{\boldsymbol{\sigma}}'_{\boldsymbol{a}})}{\psi(\boldsymbol{\sigma'_a})}\right)^{\star}, \tag{5.64}$$

where we used the shorthand notation $\boldsymbol{\sigma_a} = (\boldsymbol{\sigma}, \boldsymbol{a})$ and $\boldsymbol{\sigma'_a} = (\boldsymbol{\sigma'}, \boldsymbol{a}['])$ (the tilde does not get applied to $\boldsymbol{a}$). The probabilities are given by

$$p(\boldsymbol{\sigma_a}) = |\psi^{\star}_{\boldsymbol{v}}(\boldsymbol{\sigma},\boldsymbol{a})|^2/Z, \tag{5.65}$$

$$Z = \sum_{\boldsymbol{\sigma},\boldsymbol{a}} |\psi^{\star}_{\boldsymbol{v}}(\boldsymbol{\sigma},\boldsymbol{a})|^2. \tag{5.66}$$

Unfortunately this expression cannot be written as a squared term, so its sampled value will have worse convergence properties than something that is bounded from below such as $\mathcal{C}_{2,2}$. It is possible to perform the sampling through two independent Markov-Chains running in the purified space $\mathcal{H}_S \otimes \mathcal{H}_A$, but the increased sampling cost of this scheme could be offset by leveraging the generality of this method. As we have no constraints in the functional form of the ansatz, it would be possible to use the Convolutional ansatz presented by Sharir et Al, which allows direct sampling of a wavefunction, to efficiently sample the whole space [248]. We also remark that the Stochastic Reconfiguration method must be slightly adapted when sampling the purified space. In fact, we should not consider the metric induced by the vector bra-ket product in the space $\mathcal{H}_S \otimes \mathcal{H}_A$, but we should rather consider, as before, the distance between the resulting density matrices.

5.8 Conclusions and perspectives

In this chapter we have shown that machine learning and variational methods in quantum mechanics have much in common. We have presented a general and novel variational approach for the steady-state density matrix of open quantum many-body systems based on an artificial neural network scheme. Our method is scalable since the cost function, which ultimately measures how much the proposed state differs from the exact steady-state, can be sampled via Monte Carlo methods provided that the Liouvillian is composed of K-local operators and dissipators. We have demonstrated a proof-of-principle of this novel scheme by successfully benchmarking to brute-force finite-size simulations in the full Hilbert space for arrays of spins described by the dissipative quantum transverse Ising model. The approach that we have implemented and benchmarked is general and can be extended to deep neural networks. As a future development it would be interesting to fully generalise the structure of the ansatz, and we have already taken a first step towards this goal by sketching one possible scheme at the end of this chapter. Developments in this direction should be combined with convolutional networks that allow to greatly reduce the sampling cost while increasing the accuracy. Such approach are also capable of extracting great performance from specialised hardware such as GPUs. It would be

also interesting to study how alternative cost function behave when used in conjunction with neural-networks, as well as implementing a variational time-evolution of the density matrix or of stochastic trajectories.

CHAPTER 6

Conclusions and outlook

The framework of open quantum systems is relevant to the theoretical modelling of such systems because it can model out of equilibrium phenomena arising from the interaction with the environment or from engineered devices. The remarkable complications introduced by the coupling to the environment require advanced analytical and numerical tools to advance our understanding. We have shown that in the last decade several numerical techniques have been developed, either by generalising existing tools initially proposed for closed systems or by developing novel ideas. Nevertheless, numerical methods must be carefully chosen depending on the system and regime of interest, as they all have advantages and drawbacks. For example, semiclassical methods can be used to simulate very efficiently certain weakly interacting systems [2, 124], but fail in the quantum regime [114]. Tensor network methods provide the best performance in one dimensional systems [238], but their application to higher dimensions is not so straightforward [66]. The corner space renormalisation method faces difficulties when the state becomes highly entropic, but it treats long range correlations with ease [118, 132]. Variational methods are not limited by correlations or entropy, but by the generality and representation power of the ansatz [2, 72].

From a more fundamental point of view, within this Ph.D. thesis we have investigated the critical phenomena and the role of disorder in dissipative transitions. Disorder, in particular, is of interest because it might drive a localisation transition which could be used to engineer technological devices. With our research, we contributed to the field by developing a simple yet effective optimal sampling protocol, to efficiently compute expectation values of disordered systems through trajectory protocols [2]. This protocol is general, and can be applied to any kind of trajectory, regardless of the approximation or unravelling used to obtain it.

With regards to phase transitions, we studied the driven-dissipative Bose-Hubbard model. This fundamental model describes the coherent hopping of particles in a lattice through an exchange term, and it includes many-body features by accounting for boson-boson interaction. Incoherent single-particle losses are modelled within the Lindblad mas-

ter equation formalism. Thanks to the truncated Wigner approximation [112], we have been able to perform a finite-size scaling analysis of the model, revealing a critical slowing down of the dynamics in $2D$ lattices, suggesting that the lower critical dimension of the model is $d < 2$. We linked the slowing down, which is caused by a closing of the Liouvillian gap at the transition, to a first-order phase transition [1]. Our analysis has also revealed that in lower dimensional systems such as $1D$ chains local density fluctuations, which deplete the homogeneous mode, destroy the criticality. Interestingly, we have been able to show that critical finite-size systems exhibit lattice-level bistability of their homogeneous mode, with instantaneous switches of the whole lattice between the two phases occurring during dynamical evolution [2]. This phenomenon is not only prevented in $d < 1$, but also by any finite amount of disorder, which creates density domains. However, as long as the system is smaller than the length scale set by disorder, the system will behave qualitatively as if it was homogeneous, which might be of interest to experiments wishing to investigate this phenomenon. In the future, it would be interesting to study if coloured pumping schemes could offset the localisation induced by disorder and reinstate the criticality. We also envisage further studies on the lower critical dimension, for example by investigating fractal or frustrated lattices. Lastly, we believe that a more theoretical study of the link between bistability, long-range order and first-order dissipative transitions might uncover interesting novel results.

During our research we have often faced problems due to the sheer complexity of the simulations involved. The last part of this thesis has been devoted to creating a bridge between the language used in machine learning with that used by numerical physicists, showing that the underlying mathematical tools have a common ground. This has allowed us to develop a general method that could be used in higher dimensional ($D \geq 2$) systems and that would not be limited by the entropy [3]. We have proposed a variational principle to iteratively compute the steady-state of a Lindblad master equation. Notably, as the associated cost function is written in terms of stochastic averages, it is possible to forgo the exact evaluation, resorting to Monte Carlo sampling schemes that circumvent the otherwise exponential complexity of the Hilbert space. By combining this variational procedure with an ansatz inspired by machine learning, we have shown that it is possible to efficiently compute observables of driven-dissipative quantum systems.

The method we have proposed, while quite general, has a single constraint on the functional form of the neural network ansatz. Benchmarking on the dissipative transverse field Ising model suggested that the method could be improved by relaxing this constraint and allowing for a more general ansatz that could build more complicated correlations. We believe an interesting research direction would be to generalise the variational principle and the sampling scheme to be able to use more generic networks, and we have taken a first step in this direction by proposing the doubled purified space sampling at the end of the chapter. Completely foregoing the functional constraints will permit to use state of the art machine learning techniques such as autoregressive networks that avoid the Markov chain sampling all together. Another low-hanging fruit would be to implement symmetries into this scheme, which would further reduce the number of parameters needed while also increasing the accuracy of the sampling.

While physicists have developed several techniques to understand the effects of truncation of linear bases of the Hilbert space, it is still unclear what exactly is the limit of those

chains of nonlinear functions, and it would be extremely worthwhile to investigate this, even more so in the context of mixed states. The results, directly relevant to numerical physics, would probably also be of interest to data scientist working with ensembles of distributions.

To conclude, with the recent announcement of quantum supremacy [277] and the ongoing development of noisy intermediate scale quantum computers and simulators the development of algorithms for those devices is a very actual problem. Investigating the effectiveness of quantum neural networks, which are usually implemented in terms of parameterised quantum circuits, in encoding classical and quantum information will provide very valuable insight on future applications of those technologies. Moreover, understanding the differences between classical and quantum optimisation of quantum circuits, or the relationship between their depths and representation power would be surely worthwile.

APPENDIX A

Liouvillian superoperator as a matrix

Rewriting the Liouvillian master equation as a super-operator is quite intuitive if one explicitly writes the indices associated to the matrix-vector multiplications (assuming Einstein's summation convention and $\hat{\tilde{H}} = \hat{H}_S - i\sum_i \hat{L}_i^\dagger \hat{L}_i$):

$$(\mathcal{L}\hat{\rho})_{\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}}} = -i(\hat{\tilde{H}}\hat{\rho})_{\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}}} + i(\hat{\rho}\hat{\tilde{H}}^\dagger)_{\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}}} - \sum_i (\hat{L}_i\hat{\rho}\hat{L}_i^\dagger)_{\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}}} \tag{A.1}$$

$$= -i(\tilde{H}_{\boldsymbol{\sigma},\boldsymbol{\sigma}'}\rho_{\boldsymbol{\sigma}',\tilde{\boldsymbol{\sigma}}} + i(\rho_{\boldsymbol{\sigma}\tilde{\boldsymbol{\sigma}}'}\tilde{H}^\dagger_{\tilde{\boldsymbol{\sigma}}',\tilde{\boldsymbol{\sigma}}}) - \sum_i L_{i,\boldsymbol{\sigma},\boldsymbol{\sigma}'}\rho_{\boldsymbol{\sigma}',\tilde{\boldsymbol{\sigma}}'}L^\dagger_{i,\tilde{\boldsymbol{\sigma}}',\tilde{\boldsymbol{\sigma}}} \tag{A.2}$$

$$= \left[-i\tilde{H}_{\boldsymbol{\sigma},\boldsymbol{\sigma}'}\delta_{\tilde{\boldsymbol{\sigma}},\tilde{\boldsymbol{\sigma}}'} + i\delta_{\boldsymbol{\sigma},\boldsymbol{\sigma}'}\tilde{H}^\dagger_{\tilde{\boldsymbol{\sigma}}',\tilde{\boldsymbol{\sigma}}} - \sum_i L_{i,\boldsymbol{\sigma},\boldsymbol{\sigma}'}L^\dagger_{i,\tilde{\boldsymbol{\sigma}}',\tilde{\boldsymbol{\sigma}}}\right]\rho_{\boldsymbol{\sigma}',\tilde{\boldsymbol{\sigma}}'}. \tag{A.3}$$

Noticing that $\delta_{\tilde{\boldsymbol{\sigma}},\tilde{\boldsymbol{\sigma}}'}$ expresses the entries of the identity matrix, we can rewrite this in tensor notation, namely,

$$\mathcal{L}\hat{\rho} = \left[-i\hat{\tilde{H}} \otimes \mathbb{I} + i\,\mathbb{I} \otimes \hat{\tilde{H}}^\dagger \; - \sum_i \hat{L}_i \otimes \hat{L}_i^\dagger\right]\hat{\rho}, \tag{A.4}$$

which can be used when working with numerical solvers.

APPENDIX B

Numerics

B.1 Benchmark of the Truncated Wigner Approximation

In this appendix, we present numerical results showing that the Truncated Wigner Approximation is accurate in the regime of parameters considered in the manuscript. To do so, we compare its results to what was obtained with numerically exact methods for small systems. In Fig. B.1, we present the steady-state average population in a 4×1 array computed with the Truncated Wigner approximation and with the corner-space renormalization method [69] finding an excellent agreement between the two. The values considered are the same as in the main text. We would like to point out that for the considered value of $U/\gamma = 0.1$, a brute-force integration of the master equation for

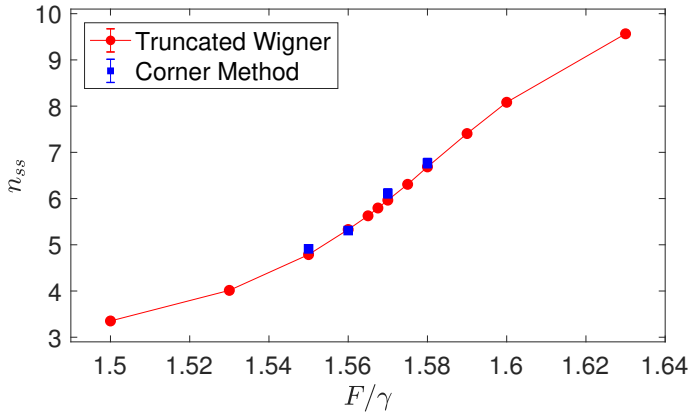

Figure B.1: Steady-state average boson occupation per site as a function of the driving F/γ in a 4×1 array: different symbols correspond to different numerical methods. The statistical error is of the order of the symbol size. Parameters are $U/\gamma = 0.1$, $zJ/\gamma = 0.9$, $\Delta/\gamma = 0.1$.

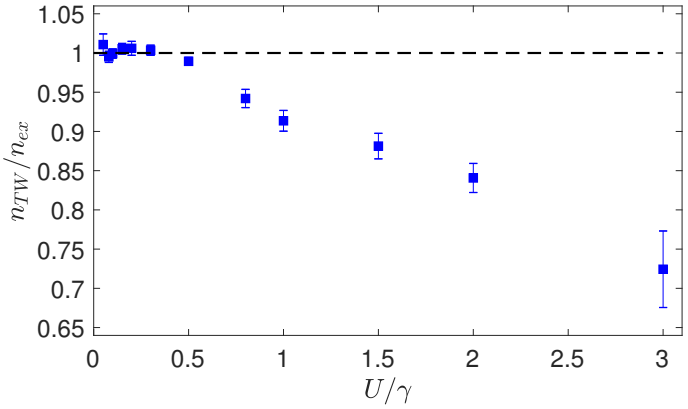

Figure B.2: Ratio between the steady-state average occupation obtained through the Truncated Wigner approximation n_{TW} and the exact Runge-Kutta Integration of the Lindblad Master Equation n_{ex} in a 2×1 array. The error bars refer to the statistical noise of the results associated to the stochastic Langevin simulations. F/γ has been varied so that $UF^2/\gamma^3 = 2.465$ is kept constant; $zJ/\gamma = 0.9$ and $\Delta/\gamma = 0.1$ are fixed. Note that the results for $U/\gamma \leq 0.2$ have been obtained with the corner-space renormalization.

a one-site system requires a cutoff of $N_{max} = 40$ bosons in order to achieve adequate numerical convergence. In a 4×1 lattice the required dimension of the Hilbert space would be $40^4 = 2.56 \cdot 10^6$ which cannot be handled numerically without more advanced methods. For the parameters considered in the main text, this lattice can still be tackled by the corner-space renormalization method (going to larger lattice sizes would require significantly larger computational resources).

In Fig. B.2 we present the ratio n_{TW}/n_{ex} between the steady-state average population obtained via the truncated Wigner approximation n_{TW} and exact methods n_{ex} as a function of the nonlinearity U/γ. We used this quantity to identify the range of values in U/γ for which the truncated Wigner approximation is quantitatively accurate, finding that for $U/\gamma \leq 0.3$ the Truncated Wigner yields results within 1% of the exact value.

APPENDIX C

Details on neural network methods for open quantum systems

C.1 Log-Derivatives of the NDM ansatz

The logarithmic derivatives of the variational density matrix $\rho_{\boldsymbol{v}}(\sigma, \tilde{\boldsymbol{\sigma}})$ appearing in Eq. (10) of the main text are computed with respect to the various components of $\boldsymbol{v}$ as follows:

$$\frac{\partial \log[\rho_{\boldsymbol{v}}(\boldsymbol{\sigma}, \tilde{\boldsymbol{\sigma}})]}{\partial \mathbf{b}_A^{(v)}} = \frac{1}{2}(\boldsymbol{\sigma} + \tilde{\boldsymbol{\sigma}}), \tag{C.1}$$

$$\frac{\partial \log[\rho_{\boldsymbol{v}}(\boldsymbol{\sigma}, \tilde{\boldsymbol{\sigma}})]}{\partial \mathbf{b}_\theta^{(v)}} = \frac{i}{2}(\boldsymbol{\sigma} - \tilde{\boldsymbol{\sigma}}), \tag{C.2}$$

$$\frac{\partial \log[\rho_{\boldsymbol{v}}(\boldsymbol{\sigma}, \tilde{\boldsymbol{\sigma}})]}{\partial b_A^{(h)[j]}} = \frac{1}{2}\left\{\frac{\mathcal{G}'}{\mathcal{G}}[\theta_A^{[j]}(\boldsymbol{\sigma})] + \frac{\mathcal{G}'}{\mathcal{G}}[\theta_A^{[j]}(\tilde{\boldsymbol{\sigma}})]\right\}, \tag{C.3}$$

$$\frac{\partial \log[\rho_{\boldsymbol{v}}(\boldsymbol{\sigma}, \tilde{\boldsymbol{\sigma}})]}{\partial b_\theta^{(h)[j]}} = \frac{i}{2}\left\{\frac{\mathcal{G}'}{\mathcal{G}}[\theta_\theta^{[j]}(\boldsymbol{\sigma})] - \frac{\mathcal{G}'}{\mathcal{G}}[\theta_\theta^{[j]}(\tilde{\boldsymbol{\sigma}})]\right\}, \tag{C.4}$$

$$\frac{\partial \log[\rho_{\boldsymbol{v}}(\boldsymbol{\sigma}, \tilde{\boldsymbol{\sigma}})]}{\partial b_A^{(a)[j]}} = \frac{\mathcal{G}'}{\mathcal{G}}\left[\frac{1}{2}(\boldsymbol{\sigma} + \tilde{\boldsymbol{\sigma}}) \cdot \mathbf{U}_A^{[j]} + \frac{i}{2}(\boldsymbol{\sigma} - \tilde{\boldsymbol{\sigma}}) \cdot \mathbf{U}_\theta^{[j]} + b_{A,j}^{(a)}\right], \tag{C.5}$$

$$\frac{\partial \log[\rho_{\boldsymbol{v}}(\boldsymbol{\sigma}, \tilde{\boldsymbol{\sigma}})]}{\partial \mathbf{W}_A^{[j]}} = \frac{1}{2}\left\{\frac{\mathcal{G}'}{\mathcal{G}}[\theta_A^{[j]}(\boldsymbol{\sigma})]\ \boldsymbol{\sigma} + \frac{\mathcal{G}'}{\mathcal{G}}[\theta_A^{[j]}(\tilde{\boldsymbol{\sigma}})]\ \tilde{\boldsymbol{\sigma}}\right\}, \tag{C.6}$$

$$\frac{\partial \log[\rho_{\boldsymbol{v}}(\boldsymbol{\sigma}, \tilde{\boldsymbol{\sigma}})]}{\partial \mathbf{W}_\theta^{[j]}} = \frac{i}{2}\left\{\frac{\mathcal{G}'}{\mathcal{G}}[\theta_\theta^{[j]}(\boldsymbol{\sigma})]\ \boldsymbol{\sigma} - \frac{\mathcal{G}'}{\mathcal{G}}[\theta_\theta^{[j]}(\tilde{\boldsymbol{\sigma}})]\ \tilde{\boldsymbol{\sigma}}\right\}, \tag{C.7}$$

$$\frac{\partial \log[\rho_{\boldsymbol{v}}(\boldsymbol{\sigma}, \tilde{\boldsymbol{\sigma}})]}{\partial \mathbf{U}_A^{[j]}} = \frac{1}{2}\ (\boldsymbol{\sigma} + \tilde{\boldsymbol{\sigma}})\ \frac{\mathcal{G}'}{\mathcal{G}}\left[\frac{1}{2}(\boldsymbol{\sigma} + \tilde{\boldsymbol{\sigma}}) \cdot \mathbf{U}_A^{[j]} + \frac{i}{2}(\boldsymbol{\sigma} - \tilde{\boldsymbol{\sigma}}) \cdot \mathbf{U}_\theta^{[j]} + b_{A,j}^{(a)}\right], \tag{C.8}$$

$$\frac{\partial \log[\rho_{\boldsymbol{v}}(\boldsymbol{\sigma}, \tilde{\boldsymbol{\sigma}})]}{\partial \mathbf{U}_\theta^{[j]}} = \frac{i}{2}\ (\boldsymbol{\sigma} - \tilde{\boldsymbol{\sigma}})\ \frac{\mathcal{G}'}{\mathcal{G}}\left[\frac{1}{2}(\boldsymbol{\sigma} + \tilde{\boldsymbol{\sigma}}) \cdot \mathbf{U}_A^{[j]} + \frac{i}{2}(\boldsymbol{\sigma} - \tilde{\boldsymbol{\sigma}}) \cdot \mathbf{U}_\theta^{[j]} + b_{A,j}^{(a)}\right], \tag{C.9}$$

where $\mathcal{G}'(x) = -e^{-x}$ and $\frac{\mathcal{G}'}{\mathcal{G}}[x] = (e^x - 1)^{-1}$.

We remark that the particular form of $\mathcal{G} = 1+e^{-x}$ used here originates from the partial trace over the hidden layer, whose neurons can assume on the binary values $\{0, 1\}$. This function, known as *softmax* in the machine learning literature, is just one of the many possible choices.

C.2 A discussion of holomorphic and non-holomoprhic function optimisation

In general an optimization problem involves finding the global minima of a real-valued scalar function $f : \mathbb{R}^N \to \mathbb{R}$. The gradient of f in $\mathbf{x} \in \mathbb{R}^N$ is defined as the column-vector $\nabla_{\mathbf{x}} f = [\partial_{x_1} f,\ \partial x_2 f, \ldots, \partial x_N f]^T$, where $\partial_{x_i} f$ is the partial derivative with respect to the i-th component of x. If the function is analytic (or, more precisely, $\mathbb{R}-$analytic), its gradient will be well defined for any value of $\mathbf{x}$. Expanding $f(\mathbf{x} + \boldsymbol{\delta})$ to linear order around $\mathbf{x}$ we obtain

$$f(\mathbf{x} + \boldsymbol{\delta}) = f(\mathbf{x}) + \boldsymbol{\delta}^T \nabla_{\mathbf{x}} f(\mathbf{x}) + \mathcal{O}(\boldsymbol{\delta}^2), \tag{C.10}$$

where $\boldsymbol{\delta}^T \nabla_{\mathbf{x}} f(\mathbf{x}) = \sum_{i=1}^N \delta_i \nabla_{x_i} f(\mathbf{x})$. Therefore, the change of the function f when updating the parameters $\mathbf{x} \to \mathbf{x} + \boldsymbol{\delta}$ is

$$|\Delta f| = f(\mathbf{x} + \boldsymbol{\delta}) - f(\mathbf{x}) = \left|\boldsymbol{\delta}^T \nabla_{\mathbf{x}} f(\mathbf{x})\right| \leq \|\boldsymbol{\delta}\| \|\nabla_{\mathbf{x}} f(\mathbf{x})\|, \tag{C.11}$$

where we used the Schwartz's inequality to bound the update. Assuming that the gradient is fixed, the update vector $\boldsymbol{\delta}$ that gives the maximum change Δf, the maximum is achieved when $\boldsymbol{\delta} = \eta \nabla_{\mathbf{x}} f(\mathbf{x})$, which is the steepest-gradient update rule.

This can be understood as the proof that any function that is $\mathbb{R}-$analytic almost everywhere[1] can be optimised with an iterative gradient descent scheme.

C.2.1 Extension to complex functions

In general a complex function $f : \mathbb{C}^N \to \mathbb{C}^M$ can be split into its real u and imaginary v part, namely

$$f(\mathbf{z}) = u(\mathbf{x}, \mathbf{y}) + iv(\mathbf{x}, \mathbf{y}), \tag{C.12}$$

where $\mathbb{C} \ni \mathbf{z} = \mathbf{x} + i\mathbf{y}$ and $\mathbf{x}, \mathbf{y} \in \mathbb{R}$. For a complex function it is possible to define the gradient and the conjugate-gradient,

$$\nabla_{\mathbf{z}} f = \nabla_{\mathbf{x}} f - i\nabla_{\mathbf{y}} f, \qquad \nabla_{\mathbf{z}^\star} f = \nabla_{\mathbf{x}} f + i\nabla_{\mathbf{y}} f. \tag{C.13}$$

Those two gradients, also known as Wirtinger gradients, can be computed operatively by assuming that $\mathbf{z}$ and $\mathbf{z}^\star$ are independent variables. A complex function is said to be holomorphic or $\mathbb{C}-$analytic if it respects the Cauchy-Riemann relation

$$\nabla_{\mathbf{z}^\star} f = 0, \tag{C.14}$$

[1] In practice, some activation functions such as rectified linear unit (relu) $relu(x) = x\theta(x)$ where θ is the Heaviside step function are non-analytic in a set of points of measure zero. As such, the probability that an iterative method will pass through that point is effectively zero, and the function can still be optimised as if it was analytic.

which means that the function has a single, well-defined gradient when treated as a function of only one parameter. In this case, one can treat the function and its derivatives like any traditional real-valued function. We remark that any operation such as $\mod[z] = \sqrt{zz^\star}$, $\mathbb{R}e[z]$ or $\mathbb{I}m[z]$ is not analytic. It is therefore easy to prove that any real-valued complex function $f : \mathbb{C}^N \to \mathbb{R}$ is non-holomorphic.

Even if a function is not $\mathbb{C}-$analytic, its real and imaginary parts u and v can still be $\mathbb{R}-$analytic, as is almost always the case in practical applications. In this case for every input $\mathbf{z}$ both the gradient and the co-gradient will be well defined. For that reason, we can still perform an iterative optimisation of a scalar real-valued complex functions can $f : \mathbb{C}^N \to \mathbb{R}$ which is $\mathbb{R}-$analytic.

To perform an iterative optimisation we need to obtain the direction of maximum change, which can be done by expanding the function to linear order in the variations $\boldsymbol{\delta}$ as

$$f(\mathbf{x}+\boldsymbol{\delta}) = f(\mathbf{x}) + \boldsymbol{\delta}^T \nabla_{\mathbf{x}} f(\mathbf{x}) + \boldsymbol{\delta}^\dagger \nabla_{\mathbf{x}^\star} f(\mathbf{x}) + \mathcal{O}(|\boldsymbol{\delta}|^2). \tag{C.15}$$

The change of the function when updating the weights will then be

$$|\Delta f| = \mathbb{R}e[\boldsymbol{\delta}^T \nabla_{\mathbf{x}^\star} f(\mathbf{x})] \leq \|\boldsymbol{\delta}\| \|\nabla_{\mathbf{x}^\star} f(\mathbf{x})\|, \tag{C.16}$$

where we used the fact that $f(\mathbf{x}) = f(\mathbf{x})^\star$. Like before, it is then trivial to see that the direction of most change is identified by $\delta = \nabla_{\mathbf{x}^\star} f(\mathbf{x})$, which proves that an iterative optimisation of a complex cost function must descend along the conjugate-gradient and not the gradient.

Chain rule

Because an $\mathbb{R}-$analytic complex function must be treated as a function of two variables, the chain rule for computing derivatives is modified accordingly. In particular, a function

$$f(\mathbf{z}) = h \circ g = h(g(\mathbf{z})), \tag{C.17}$$

has the derivatives

$$\nabla_{\mathbf{z}} f(\mathbf{z}) = \nabla_g h(g(\mathbf{x})) \nabla_{\mathbf{z}} g(\mathbf{z}) + \nabla_{g^\star} h(g(\mathbf{z})) \nabla_{\mathbf{z}} g^\star(\mathbf{x}), \tag{C.18}$$

$$\nabla_{\mathbf{z}^\star} f(\mathbf{z}) = \nabla_g h(g(\mathbf{x})) \nabla_{\mathbf{z}} g(\mathbf{z}) + \nabla_{g^\star} h(g(\mathbf{z})) \nabla_{\mathbf{z}^\star} g^\star(\mathbf{x}). \tag{C.19}$$

We remark that if the function $h(\mathbf{z})$ is holomorphic, even if g is not, then $\nabla^\star_g h(g(\mathbf{z})) = 0$ and the expressions above simplify considerably. In this case, when evaluating the conjugate-gradient of $f(\mathbf{z})$ one will only need to compute one gradient instead of two, halving the computational cost. If instead we had expressed the conjugate gradient as a sum of the gradient with respect to the real part $\nabla_{\mathrm{Re}\,\mathbf{z}}$ and the imaginary part $\nabla_{\mathrm{Im}\,\mathbf{z}}$ of $\mathbf{z}$, this would not be the case, and even if h was holomorphic two gradients would be required.

This does not make a difference when we compute the gradient analytically, because the two formulas will give the same result. However, the general approach in machine learning consists in only writing the ansatz, and letting the computer figure out automatically the gradients. This technique is called Automatic Differentiation (AD for short) [278]. In this case, knowing before-hand which functions are holomorphic and using Wirtinger derivatives will allow the program to simplify automatically the expressions.

C.3 Log-Derivatives of cost function

We will here detail the calculation of the conjugate-gradient of the cost function

$$\mathcal{C}_{2,2}(\boldsymbol{v}) = \sum_{\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}}} p_{\boldsymbol{v}}(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}}) \left| \mathcal{C}^{\mathrm{loc}}(\boldsymbol{v},\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}}) \right|^2, \tag{C.20}$$

where

$$\mathcal{C}^{\mathrm{loc}}(\boldsymbol{v},\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}}) = \sum_{\boldsymbol{\sigma}',\tilde{\boldsymbol{\sigma}}'} \mathcal{L}(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}},\boldsymbol{\sigma}',\tilde{\boldsymbol{\sigma}}') \frac{\rho_{\boldsymbol{v}}(\boldsymbol{\sigma}',\tilde{\boldsymbol{\sigma}}')}{\rho_{\boldsymbol{v}}(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}})}, \tag{C.21}$$

$$p_{\boldsymbol{v}}(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}}) = \frac{|\rho_{\boldsymbol{v}}(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}})|^2}{Z}, \tag{C.22}$$

$$Z = \mathrm{Tr}\Big[\hat{\rho}^\dagger \hat{\rho}\Big]. \tag{C.23}$$

In the following we will only assume that $\hat{\rho}_{\boldsymbol{v}}$ is $\mathbb{C}-$analytic. We recall that the log-gradients of the ansatz are given by

$$\mathcal{O}_{\boldsymbol{v}}(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}}) = \frac{\nabla_{\boldsymbol{v}} \hat{\rho}_{\boldsymbol{v}}(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}})}{\hat{\rho}_{\boldsymbol{v}}(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}})}, \qquad \mathcal{O}_{\boldsymbol{v}^\star}(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}}) = \frac{\nabla_{\boldsymbol{v}^\star} \hat{\rho}_{\boldsymbol{v}}(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}})}{\hat{\rho}_{\boldsymbol{v}}(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}})}, \tag{C.24}$$

and the latter is zero if the ansätz is holomorphic. With those definitions, we can compute the gradient of the partition function Z

$$\frac{\nabla_{\boldsymbol{v}^\star} Z}{Z} = \frac{1}{Z}\left\{ \mathrm{Tr}\Big[(\nabla_{\boldsymbol{v}^\star}\hat{\rho}^\dagger)\hat{\rho}\Big] + \mathrm{Tr}\Big[\hat{\rho}^\dagger \ \nabla_{\boldsymbol{v}^\star}\hat{\rho}\Big] \right\} \tag{C.25}$$

$$= \frac{1}{Z}\left\{ \frac{1}{Z}\mathrm{Tr}\left[\left(\frac{\nabla_{\boldsymbol{v}^\star}\hat{\rho}^\dagger}{\hat{\rho}^\dagger}\right) \ \hat{\rho}^\dagger\hat{\rho}\right] + \mathrm{Tr}\left[\hat{\rho}^\dagger\hat{\rho} \ \left(\frac{\nabla_{\boldsymbol{v}^\star}\hat{\rho}}{\hat{\rho}}\right)\right] \right\} \tag{C.26}$$

$$= \frac{1}{Z}\left\{ \mathrm{Tr}\Big[\hat{\rho}^\dagger\hat{\rho} \ \mathcal{O}_{\boldsymbol{v}^\star}\Big] + \mathrm{Tr}\Big[\hat{\rho}^\dagger\hat{\rho} \ \mathcal{O}_{\boldsymbol{v}}^\dagger\Big] \right\} \tag{C.27}$$

$$= \langle \mathcal{O}_{\boldsymbol{v}^\star}\rangle + \langle \mathcal{O}_{\boldsymbol{v}}^\dagger\rangle\,, \tag{C.28}$$

where we recall that the first term is zero if $\hat{\rho}$ is holomorphic. Similarly, one can compute the gradient of Z, namely,

$$\frac{\nabla_{\boldsymbol{v}} Z}{Z} = \langle \mathcal{O}_{\boldsymbol{v}}\rangle + \langle \mathcal{O}_{\boldsymbol{v}^\star}^\dagger\rangle\,. \tag{C.29}$$

With this, one can easily compute the gradients of the probability distribution $p_{\boldsymbol{v}}(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}})$,

$$\nabla_{\boldsymbol{v}^\star} p_{\boldsymbol{v}}(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}}) = \frac{|\rho_{\boldsymbol{v}}(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}})|^2}{Z}\Big(\mathcal{O}_{\boldsymbol{v}^\star}(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}}) + \mathcal{O}_{\boldsymbol{v}}^\dagger(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}})\Big) - \frac{|\rho_{\boldsymbol{v}}(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}})|^2}{Z}\frac{\nabla_{\boldsymbol{v}^\star} Z}{Z} \tag{C.30}$$

$$= p_{\boldsymbol{v}}(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}})\Big(\mathcal{O}_{\boldsymbol{v}^\star}(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}}) - \langle \mathcal{O}_{\boldsymbol{v}^\star}\rangle + \mathcal{O}_{\boldsymbol{v}}^\dagger(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}}) - \langle \mathcal{O}_{\boldsymbol{v}}^\dagger\rangle\Big) \tag{C.31}$$

$$= p_{\boldsymbol{v}}(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}})\Big(\Delta\mathcal{O}_{\boldsymbol{v}^\star}(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}}) + \Delta\mathcal{O}_{\boldsymbol{v}}^\dagger(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}})\Big)\,, \tag{C.32}$$

where we have defined $\Delta\mathcal{O}_{\boldsymbol{v}^\star}(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}}) = \mathcal{O}_{\boldsymbol{v}^\star}(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}}) - \langle \mathcal{O}_{\boldsymbol{v}^\star}\rangle$ and $\Delta\mathcal{O}_{\boldsymbol{v}}(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}}) = \mathcal{O}_{\boldsymbol{v}}(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}}) - \langle \mathcal{O}_{\boldsymbol{v}}\rangle$. The other gradient is

$$\nabla_{\boldsymbol{v}^\star} p_{\boldsymbol{v}}(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}}) = p_{\boldsymbol{v}}(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}})\Big(\Delta\mathcal{O}_{\boldsymbol{v}}(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}}) + \Delta\mathcal{O}_{\boldsymbol{v}^\star}^\dagger(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}})\Big)\,. \tag{C.33}$$

We also define the conjugate gradient of the local cost function as,

$$\nabla_{\boldsymbol{v}^\star}\mathcal{C}^{\mathrm{loc}}(\boldsymbol{v},\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}}) = \sum_{\boldsymbol{\sigma}',\tilde{\boldsymbol{\sigma}}'} \mathcal{L}(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}};\boldsymbol{\sigma}',\tilde{\boldsymbol{\sigma}}')\frac{\rho_{\boldsymbol{v}}(\boldsymbol{\sigma}',\tilde{\boldsymbol{\sigma}}')}{\rho_{\boldsymbol{v}}(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}})}\mathcal{O}_{\boldsymbol{v}}(\boldsymbol{\sigma}',\tilde{\boldsymbol{\sigma}}'). \tag{C.34}$$

Substituting those expressions into the definition of $\mathcal{C}_{2,2}(\boldsymbol{v})$ we obtain,

$$\nabla_{\boldsymbol{v}^\star}\mathcal{C}_{2,2}(\boldsymbol{v}) = \sum_{\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}}} \nabla_{\boldsymbol{v}^\star}p_{\boldsymbol{v}}(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}})\left|\mathcal{C}^{\mathrm{loc}}\right|^2 + p_{\boldsymbol{v}}(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}})\nabla_{\boldsymbol{v}^\star}\left|\mathcal{C}^{\mathrm{loc}}\right|^2 \tag{C.35}$$

$$= \left\langle (\Delta\mathcal{O}_{\boldsymbol{v}} + \Delta\mathcal{O}_{\boldsymbol{v}^\star})\left|\mathcal{C}^{\mathrm{loc}}\right|^2\right\rangle - \langle\Delta\mathcal{O}_{\boldsymbol{v}} + \Delta\mathcal{O}_{\boldsymbol{v}^\star}\rangle\left\langle\left|\mathcal{C}^{\mathrm{loc}}\right|^2\right\rangle, \tag{C.36}$$

which simplifies to the formula in the main text when the ansatz $\rho_{\boldsymbol{v}}$ is holomorphic and $\Delta\mathcal{O}_{\boldsymbol{v}^\star} = 0$.

C.4 Implementation details

Obtaining a performant numerical implementation of this algorithm was not an easy task. Numerical code implementing this method has been written in Julia [217] and is accessible online on GitHub [274, be sure to check the most recent version]. Work is currently underway to also implement it in the python toolbox *NetKet* [5].

The core of the algorithm requires evaluating the network and its gradient efficiently. The code performing those calculations should therefore be compiled efficiently, not interpreted; as such, it is not possible to have an efficient neural network implementation in pure Python (unless *Cython/Numba* is used). Frameworks like PyTorch are relatively easy to use and will compile the network to efficient machine code. Moreover, they provide working implementations of automatic differentiation and support GPUs out of the box. Unfortunately, existing machine learning platforms rarely support complex numbers and as such we have ruled them out. We implemented the neural network in pure Julia, which guarantees an implementation with equivalent performance to C++'s Eigen library [279], with a more concise code.

We remark that data scientists working with neural networks usually only write code defining the network and the cost function, leaving the task of generating efficient code for the gradient to Automatic Differentiation (AD) frameworks [280]. We observed that existing frameworks are engineered to compute gradients for a scalar function given a batch of inputs. Unfortunately, it is a rather complicated task to implement the liouvillian in a scalable manner that can be handled by the automatic differentiation engine. Furthermore, in 2019 support for computing the gradient of expectation values was absent in the majority of those frameworks. For that reason, we investigated using automatic differentiation to generate the gradient code, and then computing the cost function on our own, finding that AD frameworks struggle when evaluating gradients one at a time, with slowdowns in the range of $\approx 50-100$ times with respect to a hand-written kernel. We therefore resorted to handwriting them. However, we point out that due to the fast-evolving nature of software it is likely that in early 2020 several of those limitations will be overcome.

The other critical part of the implementation is the liouvillian. It is impossible to store the Liouvillian as a dense matrix, and in an extended lattice system it is also exponentially-hard to compute the sparse Liouvillian. As such, we resorted to an advanced implementation where the super-operator is stored as a sum of K-local operators, and each local term only stores the list of sites upon which it acts, and its matrix representation within that basis. The conversion between the basis on the local space and the Hilbert space of the whole system is performed on the fly.

Independent Markov chains are distributed across several computing nodes with MPI.jl and averaged on the fly.

Bibliography

[1] F. Vicentini et al. "Critical slowing down in driven-dissipative Bose-Hubbard lattices". In: *Phys. Rev. A* 97 (1 2018), p. 013853. DOI: 10.1103/PhysRevA.97.013853 (cit. on pp. vii, 4, 25, 104).

[2] F. Vicentini et al. "Optimal stochastic unraveling of disordered open quantum systems: Application to driven-dissipative photonic lattices". In: *Phys. Rev. A* 99 (3 Mar. 2019), p. 032115. DOI: 10.1103/PhysRevA.99.032115 (cit. on pp. vii, 4, 26, 40, 64, 103, 104).

[3] F. Vicentini et al. "Variational Neural-Network Ansatz for Steady States in Open Quantum Systems". In: *Phys. Rev. Lett.* 122 (25 June 2019), p. 250503. DOI: 10.1103/PhysRevLett.122.250503 (cit. on pp. vii, 3, 4, 25, 104).

[4] V. Goblot et al. "Nonlinear Polariton Fluids in a Flatband Reveal Discrete Gap Solitons". In: *Phys. Rev. Lett.* 123 (11 Sept. 2019), p. 113901. DOI: 10.1103/PhysRevLett.123.113901 (cit. on p. vii).

[5] G. Carleo et al. "NetKet: A machine learning toolkit for many-body quantum systems". In: *SoftwareX* 10 (2019), p. 100311. ISSN: 2352-7110. DOI: 10.1016/j.softx.2019.100311 (cit. on pp. vii, 115).

[6] J. P. Dowling and G. J. Milburn. "Quantum technology: the second quantum revolution". In: *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* 361.1809 (June 2003). Ed. by A. G. J. MacFarlane, pp. 1655–1674. DOI: 10.1098/rsta.2003.1227 (cit. on p. 1).

[7] C. H. Papadimitriou. *Computational complexity*. John Wiley and Sons Ltd., 2003 (cit. on p. 1).

[8] M. R. Garey, D. S. Johnson, and L. Stockmeyer. "Some simplified NP-complete problems". In: *Proceedings of the sixth annual ACM symposium on Theory of computing*. ACM. 1974, pp. 47–63 (cit. on p. 1).

[9] A. L. Blum and R. L. Rivest. "Training a 3-node neural network is NP-complete". In: *Neural Networks* 5.1 (Jan. 1992), pp. 117–127. DOI: 10.1016/s0893-6080(05)80010-3 (cit. on p. 1).

[10] P. Dagum and M. Luby. "Approximating probabilistic inference in Bayesian belief networks is NP-hard". In: *Artificial Intelligence* 60.1 (Mar. 1993), pp. 141–153. DOI: 10.1016/0004-3702(93)90036-b (cit. on p. 1).

[11] N. A. Pierce and E. Winfree. "Protein Design is NP-hard". In: *Protein Engineering, Design and Selection* 15.10 (Oct. 2002), pp. 779–782. DOI: 10.1093/protein/15.10.779 (cit. on p. 1).

[12] I. Kassal et al. "Simulating Chemistry Using Quantum Computers". In: *Annual Review of Physical Chemistry* 62.1 (May 2011), pp. 185–207. DOI: 10.1146/annurev-physchem-032210-103512 (cit. on p. 1).

[13] T. S. Cubitt, J. Eisert, and M. M. Wolf. "Extracting Dynamical Equations from Experimental Data is NP Hard". In: *Physical Review Letters* 108.12 (Mar. 2012). DOI: 10.1103/PhysRevLett.108.120503 (cit. on p. 1).

[14] P. Benioff. "The computer as a physical system: A microscopic quantum mechanical Hamiltonian model of computers as represented by Turing machines". In: *Journal of Statistical Physics* 22.5 (May 1980), pp. 563–591. DOI: 10.1007/bf01011339 (cit. on p. 1).

[15] R. P. Feynman. "Simulating physics with computers". In: *International Journal of Theoretical Physics* 21.6-7 (June 1982), pp. 467–488. DOI: 10.1007/bf02650179 (cit. on p. 1).

[16] I. M. Georgescu, S. Ashhab, and F. Nori. "Quantum simulation". In: *Rev. Mod. Phys.* 86 (1 Mar. 2014), pp. 153–185. DOI: 10.1103/RevModPhys.86.153 (cit. on p. 1).

[17] M. H. Devoret and R. J. Schoelkopf. "Superconducting Circuits for Quantum Information: An Outlook". In: *Science* 339.6124 (Mar. 2013), pp. 1169–1174. DOI: 10.1126/science.1231930 (cit. on p. 1).

[18] B. M. Terhal. "Quantum error correction for quantum memories". In: *Reviews of Modern Physics* 87.2 (Apr. 2015), pp. 307–346. DOI: 10.1103/revmodphys.87.307 (cit. on p. 2).

[19] E. Ilievski et al. "Complete Generalized Gibbs Ensembles in an Interacting Theory". In: *Phys. Rev. Lett.* 115 (15 Oct. 2015), p. 157201. DOI: 10.1103/PhysRevLett.115.157201 (cit. on p. 2).

[20] T. Langen et al. "Experimental observation of a generalized Gibbs ensemble". In: *Science* 348.6231 (2015), pp. 207–211. ISSN: 0036-8075. DOI: 10.1126/science.1257026 (cit. on p. 2).

[21] H. Breuer and F. Petruccione. *The Theory of Open Quantum Systems*. OUP Oxford, 2007. ISBN: 9780199213900. DOI: 10.1093/acprof:oso/9780199213900.001.0001 (cit. on pp. 2, 3, 6, 7, 11, 13, 15, 17, 18, 35, 53).

[22] S. Haroche and J.-M. Raimond. *Exploring the Quantum*. Oxford University Press, Aug. 2006. DOI: 10.1093/acprof:oso/9780198509141.001.0001 (cit. on pp. 2, 6).

[23] J. Q. You and F. Nori. "Atomic physics and quantum optics using superconducting circuits". In: *Nature* 474.7353 (June 2011), pp. 589–597. DOI: 10.1038/nature10122 (cit. on p. 2).

[24] A. A. Houck, H. E. Tureci, and J. Koch. "On-chip quantum simulation with superconducting circuits". In: *Nat Phys* 8 (2012), p. 292. DOI: 10.1038/nphys2251 (cit. on pp. 2, 47, 55).

[25] A. Amo et al. "Superfluidity of polaritons in semiconductor microcavities". In: *Nature Physics* 5.11 (2009), pp. 805–810. DOI: 10.1038/nphys1364 (cit. on p. 2).

[26] I. Carusotto and C. Ciuti. "Quantum fluids of light". In: *Rev. Mod. Phys.* 85 (1 Feb. 2013), pp. 299–366. DOI: 10.1103/RevModPhys.85.299 (cit. on pp. 2, 47, 49, 55, 58).

[27] M. Aspelmeyer, T. J. Kippenberg, and F. Marquardt. "Cavity optomechanics". In: *Reviews of Modern Physics* 86.4 (Dec. 2014), pp. 1391–1452. DOI: 10.1103/revmodphys.86.1391 (cit. on pp. 2, 48).

[28] C. Ciuti, P. Schwendimann, and A. Quattropani. "Theory of polariton parametric interactions in semiconductor microcavities". In: *Semiconductor Science and Technology* 18.10 (Sept. 2003), S279–S293. DOI: 10.1088/0268-1242/18/10/] (cit. on p. 2).

[29] S. Dai et al. "Tunable Phonon Polaritons in Atomically Thin van der Waals Crystals of Boron Nitride". In: *Science* 343.6175 (Mar. 2014), pp. 1125–1129. DOI: 10.1126/science.1246833 (cit. on p. 2).

[30] J. Kasprzak et al. "Bose–Einstein condensation of exciton polaritons". In: *Nature* 443.7110 (Sept. 2006), pp. 409–414. DOI: 10.1038/nature05131 (cit. on pp. 2, 48).

[31] H. Deng, H. Haug, and Y. Yamamoto. "Exciton-polariton Bose-Einstein condensation". In: *Rev. Mod. Phys.* 82 (2 May 2010), pp. 1489–1537. DOI: 10.1103/RevModPhys.82.1489 (cit. on p. 2).

[32] P. Berini and I. D. Leon. "Surface plasmon–polariton amplifiers and lasers". In: *Nature Photonics* 6.1 (Dec. 2011), pp. 16–24. DOI: 10.1038/nphoton.2011.285 (cit. on p. 2).

[33] S. D. Liberato and C. Ciuti. "Stimulated Scattering and Lasing of Intersubband Cavity Polaritons". In: *Physical Review Letters* 102.13 (Mar. 2009). DOI: 10.1103/PhysRevLett.102.136403 (cit. on p. 2).

[34] A. Collaboration. "Evidence for light-by-light scattering in heavy-ion collisions with the ATLAS detector at the LHC". In: *Nature Physics* 13.9 (Aug. 2017), pp. 852–858. DOI: 10.1038/nphys4208 (cit. on pp. 2, 49).

[35] R. J. Schoelkopf and S. M. Girvin. "Wiring up quantum systems". In: *Nature* 451.7179 (Feb. 2008), pp. 664–669. DOI: 10.1038/451664a (cit. on p. 2).

[36] M. H. Anderson et al. "Observation of Bose-Einstein Condensation in a Dilute Atomic Vapor". In: *Science* 269.5221 (July 1995), pp. 198–201. DOI: 10.1126/science.269.5221.198 (cit. on p. 2).

[37] V. V. Albert and L. Jiang. "Symmetries and conserved quantities in Lindblad master equations". In: *Phys. Rev. A* 89 (2 Feb. 2014), p. 022118. DOI: 10.1103/PhysRevA.89.022118 (cit. on pp. 3, 14, 29, 60).

[38] V. V. Albert. "Lindbladians with multiple steady states: theory and applications". PhD thesis. Yale University, 2017. arXiv: 1802.00010 [quantum-ph] (cit. on pp. 3, 14, 90).

[39] E. M. Kessler et al. "Dissipative phase transition in a central spin system". In: *Phys. Rev. A* 86 (1 July 2012), p. 012116. DOI: 10.1103/PhysRevA.86.012116 (cit. on pp. 3, 14, 47).

[40] F. Minganti et al. "Spectral theory of Liouvillians for dissipative phase transitions". In: *Phys. Rev. A* 98 (4 Oct. 2018), p. 042118. DOI: 10.1103/PhysRevA.98.042118 (cit. on pp. 3, 13, 69).

[41] S. Schütz, S. B. Jäger, and G. Morigi. "Dissipation-Assisted Prethermalization in Long-Range Interacting Atomic Ensembles". In: *Phys. Rev. Lett.* 117 (8 Aug. 2016), p. 083001. DOI: 10.1103/PhysRevLett.117.083001 (cit. on p. 3).

[42] J. Marino and S. Diehl. "Driven Markovian Quantum Criticality". In: *Phys. Rev. Lett.* 116 (7 Feb. 2016), p. 070407. DOI: 10.1103/PhysRevLett.116.070407 (cit. on p. 3).

[43] R. Rota et al. "Critical behavior of dissipative two-dimensional spin lattices". In: *Phys. Rev. B* 95 (13 Apr. 2017), p. 134431. DOI: 10.1103/PhysRevB.95.134431 (cit. on pp. 3, 33, 34, 47).

[44] O. Scarlatella, R. Fazio, and M. Schiró. "Emergent Finite Frequency Criticality of Driven-Dissipative Correlated Lattice Bosons". In: (2018). arXiv: 1805.02770 [cond-mat] (cit. on pp. 3, 52).

[45] O. Scarlatella, A. A. Clerk, and M. Schiro. "Spectral functions and negative density of states of a driven-dissipative nonlinear quantum resonator". In: *New Journal of Physics* 21.4 (Apr. 2019), p. 043040. ISSN: 1367-2630. DOI: 10.1088/1367-2630/ab0ce9 (cit. on pp. 3, 52).

[46] J. T. Young et al. "Non-equilibrium fixed points of coupled Ising models". In: (2019). arXiv: 1903.02569 [quant-ph] (cit. on pp. 3, 62, 71).

[47] E. Kapit. "The upside of noise: engineered dissipation as a resource in superconducting circuits". In: *Quantum Science and Technology* 2.3 (2017), p. 033002. DOI: 10.1088/2058-9565/aa7e5d (cit. on p. 3).

[48] M. Müller et al. "Engineered Open Systems and Quantum Simulations with Atoms and Ions". In: *Advances in Atomic, Molecular, and Optical Physics*. Ed. by P. Berman, E. Arimondo, and C. Lin. Vol. 61. Academic Press, 2012, pp. 1–80. DOI: 10.1016/B978-0-12-396482-3.00001-6 (cit. on p. 3).

[49] S. Diehl et al. "Quantum states and phases in driven open quantum systems with cold atoms". In: *Nature Physics* 4 (Sept. 2008), 878 EP -. URL: http://dx.doi.org/10.1038/nphys1073 (cit. on p. 3).

[50] A. Wolf et al. "Entangling two distant oscillators with a quantum reservoir". In: *EPL (Europhysics Letters)* 95.6 (Sept. 2011), p. 60008. DOI: 10.1209/0295-5075/95/60008 (cit. on p. 3).

[51] M. Mirrahimi et al. "Dynamically protected cat-qubits: a new paradigm for universal quantum computation". In: *New Journal of Physics* 16.4 (Apr. 2014), p. 045014. DOI: 10.1088/1367-2630/16/4/045014 (cit. on p. 3).

[52] P. D. Drummond and D. F. Walls. "Quantum theory of optical bistability. I. Nonlinear polarisability model". In: *Journal of Physics A: Mathematical and General* 13.2 (Dec. 1980), pp. 725–741. DOI: 10.1088/0305-4470/13/2/034 (cit. on pp. 3, 54, 60).

[53] N. Bartolo et al. "Exact steady state of a Kerr resonator with one- and two-photon driving and dissipation: Controllable Wigner-function multimodality and dissipative phase transitions". In: *Phys. Rev. A* 94 (3 Sept. 2016), p. 033841. DOI: 10.1103/PhysRevA.94.033841 (cit. on pp. 3, 29, 63).

[54] M. Foss-Feig et al. "Solvable Family of Driven-Dissipative Many-Body Systems". In: *Phys. Rev. Lett.* 119 (19 Nov. 2017), p. 190402. DOI: 10.1103/PhysRevLett.119.190402 (cit. on p. 3).

[55] T. Prosen. "OpenXXZSpin Chain: Nonequilibrium Steady State and a Strict Bound on Ballistic Transport". In: *Physical Review Letters* 106.21 (May 2011). DOI: 10.1103/PhysRevLett.106.217206 (cit. on p. 3).

[56] T. Prosen. "Third quantization: a general method to solve master equations for quadratic open Fermi systems". In: *New Journal of Physics* 10.4 (Apr. 2008), p. 043026. DOI: 10.1088/1367-2630/10/4/043026 (cit. on p. 3).

[57] L. M. Sieberer, M. Buchhold, and S. Diehl. "Keldysh field theory for driven open quantum systems". In: *Reports on Progress in Physics* 79.9 (2016), p. 096001. URL: http://stacks.iop.org/0034-4885/79/i=9/a=096001 (cit. on pp. 3, 25).

[58] M. F. Maghrebi and A. V. Gorshkov. "Nonequilibrium many-body steady states via Keldysh formalism". In: *Phys. Rev. B* 93 (1 Jan. 2016), p. 014307. DOI: 10.1103/PhysRevB.93.014307 (cit. on pp. 3, 25).

[59] O. Scarlatella and M. Schirò. "Quantum Impurity Models coupled to Markovian and Non Markovian Baths". In: (2019). arXiv: 1904.07679 [quant-ph] (cit. on p. 3).

[60] N. Shammah et al. "Open quantum systems with local and collective incoherent processes: Efficient numerical simulations using permutational invariance". In: *Phys. Rev. A* 98 (6 Dec. 2018), p. 063815. DOI: 10.1103/PhysRevA.98.063815 (cit. on pp. 3, 25, 29).

[61] A. Nagy and V. Savona. "Driven-dissipative quantum Monte Carlo method for open quantum systems". In: *Phys. Rev. A* 97 (5 May 2018), p. 052129. DOI: 10.1103/PhysRevA.97.052129 (cit. on pp. 3, 25).

[62] A. Biella et al. "Linked cluster expansions for open quantum systems on a lattice". In: *Phys. Rev. B* 97 (3 Jan. 2018), p. 035103. DOI: 10.1103/PhysRevB.97.035103 (cit. on pp. 3, 25, 27, 28).

[63] J. Jin et al. "Cluster Mean-Field Approach to the Steady-State Phase Diagram of Dissipative Spin Systems". In: *Phys. Rev. X* 6 (3 July 2016), p. 031011. DOI: 10.1103/PhysRevX.6.031011 (cit. on pp. 3, 25).

[64] E. Mascarenhas, H. Flayac, and V. Savona. "Matrix-product-operator approach to the nonequilibrium steady state of driven-dissipative quantum arrays". In: *Phys. Rev. A* 92 (2 Aug. 2015), p. 022116. DOI: 10.1103/PhysRevA.92.022116 (cit. on pp. 3, 25, 32).

[65] J. Cui, J. I. Cirac, and M. C. Bañuls. "Variational Matrix Product Operators for the Steady State of Dissipative Quantum Systems". In: *Physical Review Letters* 114.22 (June 2015). DOI: 10.1103/PhysRevLett.114.220601 (cit. on pp. 3, 25, 32, 85).

[66] A. Kshetrimayum, H. Weimer, and R. Orús. "A simple tensor network algorithm for two-dimensional steady states". In: *Nature Communications* 8 (1 Nov. 2017), p. 1291. DOI: 10.1038/s41467-017-01511-6 (cit. on pp. 3, 25, 31, 58, 103).

[67] D. Jaschke, S. Montangero, and L. D. Carr. "One-dimensional many-body entangled open quantum systems with tensor network methods". In: *Quantum Science and Technology* 4.1 (2019), p. 013001. URL: http://stacks.iop.org/2058-9565/4/i=1/a=013001 (cit. on pp. 3, 25).

[68] A. H. Werner et al. "Positive Tensor Network Approach for Simulating Open Quantum Many-Body Systems". In: *Phys. Rev. Lett.* 116 (23 June 2016), p. 237201. DOI: 10.1103/PhysRevLett.116.237201 (cit. on pp. 3, 25, 32).

[69] S. Finazzi et al. "Corner-Space Renormalization Method for Driven-Dissipative Two-Dimensional Correlated Systems". In: *Phys. Rev. Lett.* 115 (8 Aug. 2015), p. 080604. DOI: 10.1103/PhysRevLett.115.080604 (cit. on pp. 3, 25, 33, 34, 58, 88, 109).

[70] G. Carleo and M. Troyer. "Solving the quantum many-body problem with artificial neural networks". In: *Science* 355.6325 (Feb. 2017), pp. 602–606. DOI: 10.1126/science.aag2302 (cit. on pp. 3, 82, 91).

[71] K. Choo et al. "Symmetries and Many-Body Excitations with Neural-Network Quantum States". In: *Physical Review Letters* 121.16 (Oct. 2018). DOI: 10.1103/PhysRevLett.121.167204 (cit. on pp. 3, 91).

[72] I. Glasser et al. "Neural-Network Quantum States, String-Bond States, and Chiral Topological States". In: *Phys. Rev. X* 8 (1 Jan. 2018), p. 011006. DOI: 10.1103/PhysRevX.8.011006 (cit. on pp. 3, 103).

[73] A. Nagy and V. Savona. "Variational Quantum Monte Carlo Method with a Neural-Network Ansatz for Open Quantum Systems". In: *Physical Review Letters* 122.25 (June 2019). DOI: 10.1103/PhysRevLett.122.250501 (cit. on pp. 3, 25, 95).

[74] M. J. Hartmann and G. Carleo. "Neural-Network Approach to Dissipative Quantum Many-Body Dynamics". In: *Physical Review Letters* 122.25 (June 2019). DOI: 10.1103/PhysRevLett.122.250502 (cit. on pp. 3, 25).

[75] N. Yoshioka and R. Hamazaki. "Constructing neural stationary states for open quantum many-body systems". In: *Physical Review B* 99.21 (June 2019). DOI: 10.1103/physrevb.99.214306 (cit. on pp. 3, 25).

[76] K. Mitarai et al. "Quantum circuit learning". In: *Phys. Rev. A* 98 (3 Sept. 2018), p. 032309. DOI: 10.1103/PhysRevA.98.032309 (cit. on p. 3).

[77] W. Huggins et al. "Towards quantum machine learning with tensor networks". In: *Quantum Science and Technology* 4.2 (Jan. 2019), p. 024001. DOI: 10.1088/2058-9565/aaea94 (cit. on pp. 3, 32).

[78] V. Bergholm et al. "PennyLane: Automatic differentiation of hybrid quantum-classical computations". In: (2018). arXiv: 1811.04968 [quant-ph] (cit. on p. 3).

[79] M. Schuld et al. "Evaluating analytic gradients on quantum hardware". In: *Phys. Rev. A* 99 (3 Mar. 2019), p. 032331. DOI: 10.1103/PhysRevA.99.032331 (cit. on p. 3).

[80] M. Schuld, I. Sinayskiy, and F. Petruccione. "An introduction to quantum machine learning". In: *Contemporary Physics* 56.2 (Oct. 2014), pp. 172–185. DOI: 10.1080/00107514.2014.964942 (cit. on p. 3).

[81] J. Biamonte et al. "Quantum machine learning". In: *Nature* 549 (Sept. 2017), 195 EP -. URL: https://doi.org/10.1038/nature23474 (cit. on p. 3).

[82] H. M. Wiseman and G. J. Milburn. *Quantum Measurement and Control*. Cambridge University Press, 2010. ISBN: 9780521804424. URL: https://books.google.fr/books?id=ZNjvHaH8qA4C (cit. on pp. 6, 16–18, 35).

[83] C. Cohen-Tannoudji, B. Diu, and F. Laloë. *Quantum mechanics*. Quantum Mechanics. Wiley, 1977. ISBN: 9780471164333. URL: https://books.google.fr/books?id=rr0PAQAAMAAJ (cit. on p. 6).

[84] L. E. Ballentine. "The Statistical Interpretation of Quantum Mechanics". In: *Rev. Mod. Phys.* 42 (4 Oct. 1970), pp. 358–381. DOI: 10.1103/RevModPhys.42.358 (cit. on p. 6).

[85] C. Gardiner and P. Zoller. *Quantum Noise: A Handbook of Markovian and Non-Markovian Quantum Stochastic Methods with Applications to Quantum Optics*. Springer Series in Synergetics. Springer, 2004. ISBN: 9783540223016. URL: https://books.google.fr/books?id=a%5C_xsT8oGhdgC (cit. on pp. 7, 15, 16, 22, 44).

[86] M.-D. Choi. "Completely positive linear maps on complex matrices". In: *Linear Algebra and its Applications* 10.3 (1975), pp. 285–290. ISSN: 0024-3795. DOI: 10.1016/0024-3795(75)90075-0 (cit. on pp. 8, 31).

[87] G. Lindblad. "On the generators of quantum dynamical semigroups". In: *Communications in Mathematical Physics* 48.2 (June 1976), pp. 119–130. ISSN: 1432-0916. DOI: 10.1007/BF01608499 (cit. on p. 8).

[88] *Quantum Dynamical Semigroups and Applications*. Springer Berlin Heidelberg, 2007. DOI: 10.1007/3-540-70861-8 (cit. on p. 8).

[89] M. G. A. Paris. "The modern tools of quantum mechanics". In: *The European Physical Journal Special Topics* 203.1 (Apr. 2012), pp. 61–86. DOI: 10.1140/epjst/e2012-01535-1 (cit. on p. 11).

[90] J. Lebreuilly et al. "Stabilizing strongly correlated photon fluids with non-Markovian reservoirs". In: *Phys. Rev. A* 96 (3 Sept. 2017), p. 033828. DOI: 10.1103/PhysRevA.96.033828 (cit. on p. 12).

[91] A. Biella et al. "Phase diagram of incoherently driven strongly correlated photonic lattices". In: *Phys. Rev. A* 96 (2 Aug. 2017), p. 023839. DOI: 10.1103/PhysRevA.96.023839 (cit. on pp. 12, 33, 34, 52).

[92] D. Nigro. "On the uniqueness of the steady-state solution of the Lindblad-Gorini-Kossakowski-Sudarshan equation". In: (2018). arXiv: 1803.06279 [quant-ph] (cit. on p. 14).

[93] E. B. Davies. "Quantum stochastic processes". In: *Communications in Mathematical Physics* 15.4 (Dec. 1969), pp. 277–304. DOI: 10.1007/bf01645529 (cit. on p. 15).

[94] M. Srinivas and E. Davies. "Photon Counting Probabilities in Quantum Optics". In: *Optica Acta: International Journal of Optics* 28.7 (July 1981), pp. 981–996. DOI: 10.1080/713820643 (cit. on p. 15).

[95] H. J. Carmichael. *Statistical Methods in Quantum Optics 2*. Springer Berlin Heidelberg, 2008. DOI: 10.1007/978-3-540-71320-3 (cit. on pp. 15, 22).

[96] J. Dalibard, Y. Castin, and K. Mølmer. "Wave-function approach to dissipative processes in quantum optics". In: *Phys. Rev. Lett.* 68 (5 Feb. 1992), pp. 580–583. DOI: 10.1103/PhysRevLett.68.580 (cit. on pp. 15, 46).

[97] C. W. Gardiner. *Stochastic methods : a handbook for the natural and social sciences*. Springer, 2009, p. 447. ISBN: 9783642089626 (cit. on pp. 15, 35).

[98] P. E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*. Springer Berlin Heidelberg, 1992. DOI: 10.1007/978-3-662-12616-5 (cit. on pp. 16, 37, 58).

[99] G. E. Crooks. "Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences". In: *Physical Review E* 60.3 (Sept. 1999), pp. 2721–2726. DOI: 10.1103/physreve.60.2721 (cit. on p. 19).

[100] T. Tomé and M. J. de Oliveira. "Entropy Production in Nonequilibrium Systems at Stationary States". In: *Physical Review Letters* 108.2 (Jan. 2012). DOI: 10.1103/PhysRevLett.108.020601 (cit. on p. 19).

[101] B. Leggio et al. "Entropy production and information fluctuations along quantum trajectories". In: *Phys. Rev. A* 88 (4 Oct. 2013), p. 042111. DOI: 10.1103/PhysRevA.88.042111 (cit. on p. 19).

[102] C. Tietz et al. "Measurement of Stochastic Entropy Production". In: *Physical Review Letters* 97.5 (Aug. 2006). DOI: 10.1103/PhysRevLett.97.050602 (cit. on p. 19).

[103] P. Talkner, E. Lutz, and P. Hänggi. "Fluctuation theorems: Work is not an observable". In: *Physical Review E* 75.5 (May 2007). ISSN: 1550-2376. DOI: 10.1103/physreve.75.050102 (cit. on p. 19).

[104] S. Gherardini et al. "Reconstructing quantum entropy production to probe irreversibility and correlations". In: *Quantum Science and Technology* 3.3 (June 2018), p. 035013. DOI: 10.1088/2058-9565/aac7e1 (cit. on p. 19).

[105] J. P. Santos, G. T. Landi, and M. Paternostro. "Wigner Entropy Production Rate". In: *Physical Review Letters* 118.22 (June 2017). DOI: 10.1103/PhysRevLett.118.220601 (cit. on p. 19).

[106] S. Suomela et al. "Fluctuations of work in nearly adiabatically driven open quantum systems". In: *Physical Review E* 91.2 (Feb. 2015). ISSN: 1550-2376. DOI: 10.1103/physreve.91.022126 (cit. on p. 19).

[107] C. Elouard et al. "The role of quantum measurement in stochastic thermodynamics". In: *npj Quantum Information* 3.1 (2017), p. 9. DOI: 10.1038/s41534-017-0008-4 (cit. on p. 19).

[108] J. M. Horowitz and J. M. R. Parrondo. "Entropy production along nonequilibrium quantum jump trajectories". In: *New Journal of Physics* 15.8 (Aug. 2013), p. 085028. DOI: 10.1088/1367-2630/15/8/085028 (cit. on p. 19).

[109] N. Cottet et al. "Observing a quantum Maxwell demon at work". In: *Proceedings of the National Academy of Sciences* 114.29 (July 2017), pp. 7561–7564. DOI: 10.1073/pnas.1704827114 (cit. on p. 19).

[110] K. E. Cahill and R. J. Glauber. "Density Operators and Quasiprobability Distributions". In: *Phys. Rev.* 177 (5 Jan. 1969), pp. 1882–1902. DOI: 10.1103/PhysRev.177.1882 (cit. on p. 21).

[111] H. J. Carmichael. *Statistical Methods in Quantum Optics 1: Master Equations and Fokker-Planck Equations*. Physics and Astronomy Online Library. Springer, 1998. ISBN: 9783540548829. URL: https://books.google.fr/books?id=ocgRgM-yJacC (cit. on pp. 22, 35, 58).

[112] K. Vogel and H. Risken. "Quasiprobability distributions in dispersive optical bistability". In: *Phys. Rev. A* 39 (9 May 1989), pp. 4675–4683. DOI: 10.1103/PhysRevA.39.4675 (cit. on pp. 22, 37, 58, 60, 104).

[113] H. Risken. *The Fokker-Planck Equation*. Springer Berlin Heidelberg, 1989. DOI: 10.1007/978-3-642-61544-3 (cit. on pp. 22, 38, 39).

[114] K. Vogel and H. Risken. "Quantum-tunneling rates and stationary solutions in dispersive optical bistability". In: *Phys. Rev. A* 38 (5 Sept. 1988), pp. 2409–2422. DOI: 10.1103/PhysRevA.38.2409 (cit. on pp. 22, 39, 103).

[115] P. Deuar and P. D. Drummond. "GaugePrepresentations for quantum-dynamical problems: Removal of boundary terms". In: *Physical Review A* 66.3 (Sept. 2002). DOI: 10.1103/physreva.66.033812 (cit. on pp. 22, 58).

[116] P. Deuar. "First-principles quantum simulations of many-mode open interacting Bose gases using stochastic gauge methods". PhD thesis. University of Queensland, June 2005. arXiv: 0507023 [cond-mat] (cit. on pp. 22, 39, 58).

[117] H. Aoki et al. "Nonequilibrium dynamical mean-field theory and its applications". In: *Reviews of Modern Physics* 86.2 (June 2014), pp. 779–837. DOI: 10.1103/revmodphys.86.779 (cit. on p. 25).

[118] W. Casteels et al. "Truncated correlation hierarchy schemes for driven-dissipative multimode quantum systems". In: *New Journal of Physics* 18.9 (Sept. 2016), p. 093007. DOI: 10.1088/1367-2630/18/9/093007 (cit. on pp. 25, 103).

[119] T. Grujic et al. "Non-equilibrium many-body effects in driven nonlinear resonator arrays". In: *New Journal of Physics* 14.10 (2012), p. 103025. URL: http://stacks.iop.org/1367-2630/14/i=10/a=103025 (cit. on p. 25).

[120] D. Huybrechts and M. Wouters. "Cluster methods for the description of a driven-dissipative spin model". In: *Phys. Rev. A* 99 (4 Apr. 2019), p. 043841. DOI: 10.1103/PhysRevA.99.043841 (cit. on pp. 25, 29).

[121] W. Casteels, R. M. Wilson, and M. Wouters. "Gutzwiller Monte Carlo approach for a critical dissipative spin model". In: *Physical Review A* 97.6 (June 2018). DOI: 10.1103/physreva.97.062107 (cit. on pp. 25, 29).

[122] E. Mascarenhas. "Diffusive-Gutzwiller approach to the quadratically driven photonic lattice". In: (2017). arXiv: 1712.00987 [quant-ph] (cit. on pp. 25, 29).

[123] S. B. Jäger et al. "Semiclassical theory of synchronization-assisted cooling". In: *Phys. Rev. A* 95 (6 June 2017), p. 063852. DOI: 10.1103/PhysRevA.95.063852 (cit. on p. 25).

[124] W. Verstraelen and M. Wouters. "Gaussian Quantum Trajectories for the Variational Simulation of Open Quantum-Optical Systems". In: *Applied Sciences* 8.9 (Aug. 2018), p. 1427. DOI: 10.3390/app8091427 (cit. on pp. 25, 39, 103).

[125] H. Weimer. "Variational principle for steady states of dissipative quantum many-body systems". In: *Physical Review Letters* 114.4 (2015), pp. 1–6. ISSN: 10797114. DOI: 10.1103/PhysRevLett.114.040402. eprint: 1409.8307 (cit. on pp. 25, 88, 100).

[126] D. J. Higham and L. N. Trefethen. "Stiffness of odes". In: *BIT Numerical Mathematics* 33.2 (1993), pp. 285–303. DOI: 10.1007/BF01989751 (cit. on p. 26).

[127] J. Butcher. "A history of Runge-Kutta methods". In: *Applied Numerical Mathematics* 20.3 (Mar. 1996), pp. 247–260. DOI: 10.1016/0168-9274(95)00108-5 (cit. on p. 26).

[128] J. Dormand and P. Prince. "A family of embedded Runge-Kutta formulae". In: *Journal of Computational and Applied Mathematics* 6.1 (Mar. 1980), pp. 19–26. DOI: 10.1016/0771-050x(80)90013-3 (cit. on p. 27).

[129] C. Tsitouras. "Runge–Kutta pairs of order 5(4) satisfying only the first column simplifying assumption". In: *Computers & Mathematics with Applications* 62.2 (July 2011), pp. 770–775. DOI: 10.1016/j.camwa.2011.06.002 (cit. on p. 27).

[130] M. C. Gutzwiller. "Effect of Correlation on the Ferromagnetism of Transition Metals". In: *Physical Review Letters* 10.5 (Mar. 1963), pp. 159–162. DOI: 10.1103/PhysRevLett.10.159 (cit. on p. 27).

[131] V. Savona. "Spontaneous symmetry breaking in a quadratically driven nonlinear photonic lattice". In: *Phys. Rev. A* 96 (3 Sept. 2017), p. 033826. DOI: 10.1103/PhysRevA.96.033826 (cit. on p. 29).

[132] R. Rota et al. "Quantum Critical Regime in a Quadratically Driven Nonlinear Photonic Lattice". In: *Phys. Rev. Lett.* 122 (11 Mar. 2019), p. 110405. DOI: 10.1103/PhysRevLett.122.110405 (cit. on pp. 29, 30, 33, 34, 103).

[133] H. A. van der Vorst. "Bi-CGSTAB: A Fast and Smoothly Converging Variant of Bi-CG for the Solution of Nonsymmetric Linear Systems". In: *SIAM Journal on Scientific and Statistical Computing* 13.2 (Mar. 1992), pp. 631–644. DOI: 10.1137/0913035 (cit. on p. 30).

[134] M. H. Gutknecht. "Variants of BICGSTAB for Matrices with Complex Spectrum". In: *SIAM Journal on Scientific Computing* 14.5 (Sept. 1993), pp. 1020–1033. DOI: 10.1137/0914062 (cit. on p. 30).

[135] F. Verstraete, J. J. Garcia-Ripoll, and J. I. Cirac. "Matrix Product Density Operators: Simulation of Finite-Temperature and Dissipative Systems". In: *Phys. Rev. Lett.* 93 (20 Nov. 2004), p. 207204. DOI: 10.1103/PhysRevLett.93.207204 (cit. on p. 31).

[136] M. Kliesch, D. Gross, and J. Eisert. "Matrix-Product Operators and States: NP-Hardness and Undecidability". In: *Phys. Rev. Lett.* 113 (16 Oct. 2014), p. 160503. DOI: 10.1103/PhysRevLett.113.160503 (cit. on p. 31).

[137] M. Zwolak and G. Vidal. "Mixed-State Dynamics in One-Dimensional Quantum Lattice Systems: A Time-Dependent Superoperator Renormalization Algorithm". In: *Phys. Rev. Lett.* 93 (20 Nov. 2004), p. 207205. DOI: 10.1103/PhysRevLett.93.207205 (cit. on p. 31).

[138] G. D. las Cuevas et al. "Purifications of multipartite states: limitations and constructive methods". In: *New Journal of Physics* 15.12 (Dec. 2013), p. 123021. DOI: 10.1088/1367-2630/15/12/123021 (cit. on p. 32).

[139] G. De las Cuevas et al. "Fundamental limitations in the purifications of tensor networks". In: *Journal of Mathematical Physics* 57.7 (July 2016), p. 071902. ISSN: 1089-7658. DOI: 10.1063/1.4954983 (cit. on p. 32).

[140] E. M. Stoudenmire. "Learning relevant features of data with multi-scale tensor networks". In: *Quantum Science and Technology* 3.3 (Apr. 2018), p. 034003. DOI: 10.1088/2058-9565/aaba1a (cit. on p. 32).

[141] S. Cheng et al. "Tree tensor networks for generative modeling". In: *Physical Review B* 99.15 (Apr. 2019). DOI: 10.1103/physrevb.99.155131 (cit. on p. 32).

[142] W. Casteels et al. "Probing photon correlations in the dark sites of geometrically frustrated cavity lattices". In: *Phys. Rev. A* 93 (4 Apr. 2016), p. 043833. DOI: 10.1103/PhysRevA.93.043833 (cit. on p. 33).

[143] Y.-Y. Shi, L.-M. Duan, and G. Vidal. "Classical simulation of quantum many-body systems with a tree tensor network". In: *Phys. Rev. A* 74 (2 Aug. 2006), p. 022320. DOI: 10.1103/PhysRevA.74.022320 (cit. on p. 34).

[144] R. Rota et al. "Dynamical properties of dissipative XYZ Heisenberg lattices". In: *New Journal of Physics* 20.4 (2018), p. 045003. DOI: 10.1088/1367-2630/aab703 (cit. on p. 34).

[145] M. B. Plenio and P. L. Knight. "The quantum-jump approach to dissipative dynamics in quantum optics". In: *Rev. Mod. Phys.* 70 (1 Jan. 1998), pp. 101–144. DOI: 10.1103/RevModPhys.70.101 (cit. on p. 34).

[146] J. Gambetta et al. "Quantum trajectory approach to circuit QED: Quantum jumps and the Zeno effect". In: *Phys. Rev. A* 77 (1 Jan. 2008), p. 012112. DOI: 10.1103/PhysRevA.77.012112 (cit. on p. 34).

[147] L. Accardi, A. Frigerio, and J. T. Lewis. "Quantum stochastic processes". In: *Publications of the Research Institute for Mathematical Sciences* 18.1 (1982), pp. 97–133 (cit. on p. 35).

[148] A. J. Daley. "Quantum trajectories and open many-body quantum systems". In: *Advances in Physics* 63.2 (2014), pp. 77–149. DOI: 10.1080/00018732.2014.933502 (cit. on pp. 35, 42, 58).

[149] M. J. Gander and S. Vandewalle. "Analysis of the Parareal Time-Parallel Time-Integration Method". In: *SIAM Journal on Scientific Computing* 29.2 (Jan. 2007), pp. 556–578. DOI: 10.1137/05064607x (cit. on p. 35).

[150] R. Speck and D. Ruprecht. "Toward fault-tolerant parallel-in-time integration with PFASST". In: *Parallel Computing* 62 (2017), pp. 20–37. ISSN: 0167-8191. DOI: https://doi.org/10.1016/j.parco.2016.12.001 (cit. on p. 35).

[151] K. Mølmer, Y. Castin, and J. Dalibard. "Monte Carlo wave-function method in quantum optics". In: *Josa B* 10.3 (1993), pp. 524–538. ISSN: 1520-8540. URL: http://www.opticsinfobase.org/abstract.cfm?uri=josab-10-3-524 (cit. on pp. 36, 46).

[152] J. Johansson, P. Nation, and F. Nori. "QuTiP: An open-source Python framework for the dynamics of open quantum systems". In: *Computer Physics Communications* 183.8 (Aug. 2012), pp. 1760–1772. DOI: 10.1016/j.cpc.2012.02.021 (cit. on pp. 36, 45).

[153] J. Johansson, P. Nation, and F. Nori. "QuTiP 2: A Python framework for the dynamics of open quantum systems". In: *Computer Physics Communications* 184.4 (Apr. 2013), pp. 1234–1240. DOI: 10.1016/j.cpc.2012.11.019 (cit. on pp. 36, 45).

[154] S. Krämer et al. "QuantumOptics.jl: A Julia framework for simulating open quantum systems". In: *Computer Physics Communications* 227 (June 2018), pp. 109–116. DOI: 10.1016/j.cpc.2018.02.004 (cit. on pp. 36, 45).

[155] X. Han and P. E. Kloeden. *Random Ordinary Differential Equations and Their Numerical Solution*. Springer Singapore, 2017. DOI: 10.1007/978-981-10-6265-0 (cit. on p. 36).

[156] Y. Asai. "Numerical methods for random ordinary differential equations and their applications in biology and medicine". PhD thesis. 2016, p. 178. URL: http://publikationen.ub.uni-frankfurt.de/frontdoor/index/index/docId/40146 (cit. on p. 36).

[157] M. Sawerwain and J. Wiśniewska. "QTM: Computational package using MPI protocol for Quantum Trajectories Method". In: *PLOS ONE* 13.12 (Dec. 2018). Ed. by G. Juzeliunas, e0208263. DOI: 10.1371/journal.pone.0208263 (cit. on p. 36).

[158] I. Karatzas and S. E. Shreve. "Brownian Motion". In: *Brownian Motion and Stochastic Calculus*. Springer New York, 1998, pp. 47–127. DOI: 10.1007/978-1-4612-0949-2_2 (cit. on p. 37).

[159] A. Rößler. "Runge–Kutta Methods for the Strong Approximation of Solutions of Stochastic Differential Equations". In: *SIAM Journal on Numerical Analysis* 48.3 (Jan. 2010), pp. 922–952. DOI: 10.1137/09076636x (cit. on p. 37).

[160] C. Rackauckas and Q. Nie. "Stability-Optimized High Order Methods and Stiffness Detection for Pathwise Stiff Stochastic Differential Equations". In: (2018). arXiv: 1804.04344 [math] (cit. on pp. 37, 58).

[161] C. Rackauckas and Q. Nie. "Adaptive methods for stochastic differential equations via natural embeddings and rejection sampling with memory". In: *Discrete and Continuous Dynamical Systems - Series B* 22.7 (2017), pp. 2731–2761. ISSN: 1531-3492. DOI: 10.3934/dcdsb.2017133 (cit. on pp. 37, 58).

[162] L. Bonnes and A. M. Läuchli. "Superoperators vs. Trajectories for Matrix Product State Simulations of Open Quantum System: A Case Study". In: (2014). arXiv: 1411.4831 [cond-mat] (cit. on p. 37).

[163] C. Yang et al. "Matrix Product States for Quantum Stochastic Modeling". In: *Physical Review Letters* 121.26 (Dec. 2018). DOI: 10.1103/PhysRevLett.121.260602 (cit. on p. 37).

[164] K. Temme and F. Verstraete. "Stochastic Matrix Product States". In: *Physical Review Letters* 104.21 (May 2010). DOI: 10.1103/PhysRevLett.104.210502 (cit. on p. 37).

[165] D. M. Basko, I. L. Aleiner, and B. L. Altshuler. "Metal–insulator transition in a weakly interacting many-electron system with localized single-particle states". In: *Annals of Physics* 321.5 (2006), pp. 1126–1205. ISSN: 0003-4916. DOI: 10.1016/j.aop.2005.11.014 (cit. on p. 40).

[166] V. Oganesyan and D. A. Huse. "Localization of interacting fermions at high temperature". In: *Phys. Rev. B* 75 (15 Apr. 2007), p. 155111. DOI: 10.1103/PhysRevB.75.155111 (cit. on p. 40).

[167] R. Nandkishore and D. A. Huse. "Many-Body Localization and Thermalization in Quantum Statistical Mechanics". In: *Annual Review of Condensed Matter Physics* 6.1 (2015), pp. 15–38. DOI: 10.1146/annurev-conmatphys-031214-014726 (cit. on p. 40).

[168] X. Xu, C. Guo, and D. Poletti. "Interplay of interaction and disorder in the steady state of an open quantum system". In: *Phys. Rev. B* 97 (14 Apr. 2018), p. 140201. DOI: 10.1103/PhysRevB.97.140201 (cit. on p. 40).

[169] I. Vakulchyk et al. "Signatures of many-body localization in steady states of open quantum systems". In: *Phys. Rev. B* 98 (2 July 2018), p. 020202. DOI: 10.1103/PhysRevB.98.020202 (cit. on p. 40).

[170] C. E. Máximo et al. "Spatial and temporal localization of light in two dimensions". In: *Phys. Rev. A* 92 (6 Dec. 2015), p. 062702. DOI: 10.1103/PhysRevA.92.062702 (cit. on p. 40).

[171] G. L. Celardo, M. Angeli, and R. Kaiser. "Localization of light in subradiant Dicke states: a mobility edge in the imaginary axis". In: (2017). arXiv: 1702.04506 [cond-mat] (cit. on p. 40).

[172] A. Biella et al. "Subradiant hybrid states in the open 3D Anderson-Dicke model". In: *EPL (Europhysics Letters)* 103.5 (2013), p. 57009. URL: http://stacks.iop.org/0295-5075/103/i=5/a=57009 (cit. on p. 40).

[173] G. Celardo et al. "Interplay of superradiance and disorder in the Anderson Model". In: *Fortschritte der Physik* 61.2-3 (2012), pp. 250–260. DOI: 10.1002/prop.201200082 (cit. on p. 40).

[174] T. F. See et al. "Strongly correlated photon transport in nonlinear photonic lattice with disorder: Probing signatures of the localization transition". In: (2018). arXiv: 1807.07882 [quant-ph] (cit. on p. 40).

[175] M. F. Maghrebi. "Fragile fate of driven-dissipative XY phase in two dimensions". In: *Phys. Rev. B* 96 (17 Nov. 2017), p. 174304. DOI: 10.1103/PhysRevB.96.174304 (cit. on p. 40).

[176] G. Kulaitis et al. "Disordered driven coupled cavity arrays: Nonequilibrium stochastic mean-field theory". In: *Physical Review A* 87.1 (2013), p. 013840. DOI: 10.1103/PhysRevA.87.013840 (cit. on p. 40).

[177] J. Kenney and E. Keeping. *Mathematics of statistics*. Mathematics of Statistics 1. Van Nostrand, 1947. URL: https://archive.org/details/MathematicsOfStatisticsPartI/page/n31 (cit. on p. 43).

[178] S. Sachdev. *Quantum Phase Transitions*. Cambridge University Press, 2001. ISBN: 9780521004541. URL: https://books.google.fr/books?id=Ih%5C_E05N5TZQC (cit. on pp. 47, 68).

[179] H. J. Carmichael. "Breakdown of Photon Blockade: A Dissipative Quantum Phase Transition in Zero Dimensions". In: *Phys. Rev. X* 5 (3 Sept. 2015), p. 031028. DOI: 10.1103/PhysRevX.5.031028 (cit. on p. 47).

[180] J. J. Mendoza-Arenas et al. "Beyond mean-field bistability in driven-dissipative lattices: Bunching-antibunching transition and quantum simulation". In: *Phys. Rev. A* 93 (2 Feb. 2016), p. 023821. DOI: 10.1103/PhysRevA.93.023821 (cit. on p. 47).

[181] W. Casteels et al. "Power laws in the dynamic hysteresis of quantum nonlinear photonic resonators". In: *Phys. Rev. A* 93 (3 Mar. 2016), p. 033824. DOI: 10.1103/PhysRevA.93.033824 (cit. on pp. 47, 55, 60).

[182] W. Casteels and C. Ciuti. "Quantum entanglement in the spatial-symmetry-breaking phase transition of a driven-dissipative Bose-Hubbard dimer". In: *Phys. Rev. A* 95 (1 Jan. 2017), p. 013812. DOI: 10.1103/PhysRevA.95.013812 (cit. on p. 47).

[183] W. Casteels, R. Fazio, and C. Ciuti. "Critical dynamical properties of a first-order dissipative phase transition". In: *Phys. Rev. A* 95 (1 Jan. 2017), p. 012128. DOI: 10.1103/PhysRevA.95.012128 (cit. on pp. 47, 55–57, 61–63).

[184] L. M. Sieberer et al. "Dynamical Critical Phenomena in Driven-Dissipative Systems". In: *Phys. Rev. Lett.* 110 (19 May 2013), p. 195301. DOI: 10.1103/PhysRevLett.110.195301 (cit. on p. 47).

[185] L. M. Sieberer et al. "Nonequilibrium functional renormalization for driven-dissipative Bose-Einstein condensation". In: *Phys. Rev. B* 89 (13 Apr. 2014), p. 134310. DOI: 10.1103/PhysRevB.89.134310 (cit. on p. 47).

[186] E. Altman et al. "Two-Dimensional Superfluidity of Exciton Polaritons Requires Strong Anisotropy". In: *Phys. Rev. X* 5 (1 Feb. 2015), p. 011017. DOI: 10.1103/PhysRevX.5.011017 (cit. on p. 47).

[187] T. E. Lee, S. Gopalakrishnan, and M. D. Lukin. "Unconventional Magnetism via Optical Pumping of Interacting Spin Systems". In: *Phys. Rev. Lett.* 110 (25 June 2013), p. 257204. DOI: 10.1103/PhysRevLett.110.257204 (cit. on pp. 47, 48).

[188] C. Noh and D. G. Angelakis. "Quantum simulations and many-body physics with light". In: *Reports on Progress in Physics* 80.1 (2017), p. 016401. DOI: 10.1088/0034-4885/80/1/016401 (cit. on p. 47).

[189] M. J. Hartmann. "Quantum simulation with interacting photons". In: *Journal of Optics* 18.10 (2016), p. 104005. DOI: 10.1088/2040-8978/18/10/104005 (cit. on p. 47).

[190] A. Amo and J. Bloch. "Exciton-polaritons in lattices: A non-linear photonic simulator". In: *Comptes Rendus Physique* 17.8 (2016). Polariton physics / Physique des polaritons, pp. 934–945. ISSN: 1631-0705. DOI: 10.1016/j.crhy.2016.08.007 (cit. on p. 47).

[191] T. Jacqmin et al. "Direct Observation of Dirac Cones and a Flatband in a Honeycomb Lattice for Polaritons". In: *Phys. Rev. Lett.* 112 (11 Mar. 2014), p. 116402. DOI: 10.1103/PhysRevLett.112.116402 (cit. on p. 47).

[192] M. Fitzpatrick et al. "Observation of a Dissipative Phase Transition in a One-Dimensional Circuit QED Lattice". In: *Phys. Rev. X* 7 (1 Feb. 2017), p. 011016. DOI: 10.1103/PhysRevX.7.011016 (cit. on pp. 47, 51, 55, 61).

[193] R. Labouvie et al. "Bistability in a Driven-Dissipative Superfluid". In: *Phys. Rev. Lett.* 116 (23 June 2016), p. 235302. DOI: 10.1103/PhysRevLett.116.235302 (cit. on pp. 47, 48, 51).

[194] M. Saffman, T. G. Walker, and K. Mølmer. "Quantum information with Rydberg atoms". In: *Rev. Mod. Phys.* 82 (3 Aug. 2010), pp. 2313–2363. DOI: 10.1103/RevModPhys.82.2313 (cit. on p. 48).

[195] M. Viteau et al. "Cooperative Excitation and Many-Body Interactions in a Cold Rydberg Gas". In: *Phys. Rev. Lett.* 109 (5 July 2012), p. 053002. DOI: 10.1103/PhysRevLett.109.053002 (cit. on p. 48).

[196] J. J. Hopfield. "Theory of the Contribution of Excitons to the Complex Dielectric Constant of Crystals". In: *Phys. Rev.* 112 (5 Dec. 1958), pp. 1555–1567. DOI: 10.1103/PhysRev.112.1555 (cit. on p. 50).

[197] S. R. K. Rodriguez et al. "Probing a Dissipative Phase Transition via Dynamical Optical Hysteresis". In: *Phys. Rev. Lett.* 118 (24 June 2017), p. 247402. DOI: 10.1103/PhysRevLett.118.247402 (cit. on pp. 50, 55, 60, 61).

[198] N. C. Zambon et al. "Optically controlling the emission chirality of microlasers". In: *Nature Photonics* 13.4 (Mar. 2019), pp. 283–288. DOI: 10.1038/s41566-019-0380-z (cit. on pp. 50, 51).

[199] R. Ma et al. "A dissipatively stabilized Mott insulator of photons". In: *Nature* 566.7742 (Feb. 2019), pp. 51–57. DOI: 10.1038/s41586-019-0897-9 (cit. on p. 52).

[200] J. K. Freericks and H. Monien. "Phase diagram of the Bose-Hubbard Model". In: *Europhysics Letters (EPL)* 26.7 (June 1994), pp. 545–550. DOI: 10.1209/0295-5075/26/7/012 (cit. on p. 54).

[201] D. Jaksch et al. "Cold Bosonic Atoms in Optical Lattices". In: *Phys. Rev. Lett.* 81 (15 Oct. 1998), pp. 3108–3111. DOI: 10.1103/PhysRevLett.81.3108 (cit. on p. 54).

[202] B. Damski et al. "Atomic Bose and Anderson Glasses in Optical Lattices". In: *Phys. Rev. Lett.* 91 (8 Aug. 2003), p. 080403. DOI: 10.1103/PhysRevLett.91.080403 (cit. on p. 54).

[203] D. Jaksch and P. Zoller. "The cold atom Hubbard toolbox". In: *Annals of Physics* 315.1 (Jan. 2005), pp. 52–79. DOI: 10.1016/j.aop.2004.09.010 (cit. on p. 54).

[204] V. W. Scarola and S. Das Sarma. "Quantum Phases of the Extended Bose-Hubbard Hamiltonian: Possibility of a Supersolid State of Cold Atoms in Optical Lattices". In: *Phys. Rev. Lett.* 95 (3 July 2005), p. 033003. DOI: 10.1103/PhysRevLett.95.033003 (cit. on p. 54).

[205] B. Capogrosso-Sansone et al. "Monte Carlo study of the two-dimensional Bose-Hubbard model". In: *Phys. Rev. A* 77 (1 Jan. 2008), p. 015602. DOI: 10.1103/PhysRevA.77.015602 (cit. on p. 54).

[206] M. Greiner et al. "Quantum phase transition from a superfluid to a Mott insulator in a gas of ultracold atoms". In: *Nature* 415.6867 (Jan. 2002), pp. 39–44. DOI: 10.1038/415039a (cit. on p. 54).

[207] K. Winkler et al. "Repulsively bound atom pairs in an optical lattice". In: *Nature* 441.7095 (June 2006), pp. 853–856. DOI: 10.1038/nature04918 (cit. on p. 54).

[208] M. J. Hartmann, F. G. S. L. Brandao, and M. B. Plenio. "Strongly interacting polaritons in coupled arrays of cavities". In: *Nat Phys* 2.12 (Dec. 2006), pp. 849–855. ISSN: 1745-2473. DOI: 10.1038/nphys462 (cit. on p. 55).

[209] A. D. Greentree et al. "Quantum phase transitions of light". In: *Nature Physics* 2.12 (2006), pp. 856–861. DOI: 10.1038/nphys466 (cit. on p. 55).

[210] D. G. Angelakis, M. F. Santos, and S. Bose. "Photon-blockade-induced Mott transitions and XY spin models in coupled cavity arrays". In: *Phys. Rev. A* 76 (3 Sept. 2007), p. 031805. DOI: 10.1103/PhysRevA.76.031805 (cit. on p. 55).

[211] A. Tomadin et al. "Signatures of the superfluid-insulator phase transition in laser-driven dissipative nonlinear cavity arrays". In: *Phys. Rev. A* 81 (6 June 2010), p. 061801. DOI: 10.1103/PhysRevA.81.061801 (cit. on p. 55).

[212] C.-W. Wu et al. "Quantum phase transition of light in a one-dimensional photon-hopping-controllable resonator array". In: *Phys. Rev. A* 84 (4 Oct. 2011), p. 043827. DOI: 10.1103/PhysRevA.84.043827 (cit. on p. 55).

[213] M. Biondi et al. "Nonequilibrium gas-liquid transition in the driven-dissipative photonic lattice". In: *Phys. Rev. A* 96 (4 Oct. 2017), p. 043809. DOI: 10.1103/PhysRevA.96.043809 (cit. on p. 55).

[214] A. Le Boité, G. Orso, and C. Ciuti. "Steady-State Phases and Tunneling-Induced Instabilities in the Driven Dissipative Bose-Hubbard Model". In: *Phys. Rev. Lett.* 110 (23 June 2013), p. 233601. DOI: 10.1103/PhysRevLett.110.233601 (cit. on pp. 55, 56, 60).

[215] A. Le Boité, G. Orso, and C. Ciuti. "Bose-Hubbard model: Relation between driven-dissipative steady states and equilibrium quantum phases". In: *Phys. Rev. A* 90 (6 Dec. 2014), p. 063821. DOI: 10.1103/PhysRevA.90.063821 (cit. on pp. 55, 60).

[216] M. Foss-Feig et al. "Emergent equilibrium in many-body optical bistability". In: *Phys. Rev. A* 95 (4 2017), p. 043826. DOI: 10.1103/PhysRevA.95.043826 (cit. on pp. 55, 60, 62, 71).

[217] J. Bezanson et al. "Julia: A Fresh Approach to Numerical Computing". In: *SIAM Review* 59.1 (2017), pp. 65–98. DOI: 10.1137/141000671 (cit. on pp. 58, 115).

[218] C. Rackauckas and Q. Nie. "DifferentialEquations.jl, A Performant and Feature-Rich Ecosystem for Solving Differential Equations in Julia". In: *Journal of Open Research Software* 5 (2017), p. 15. DOI: 10.5334/jors.151 (cit. on p. 58).

[219] R. M. Wilson et al. "Collective phases of strongly interacting cavity photons". In: *Phys. Rev. A* 94 (3 Sept. 2016), p. 033801. DOI: 10.1103/PhysRevA.94.033801 (cit. on p. 60).

[220] T. Fink et al. "Signatures of a dissipative phase transition in photon correlation measurements". In: *Nature Physics* (Dec. 2017). DOI: 10.1038/s41567-017-0020-9 (cit. on pp. 61, 62).

[221] B. S. Chissom. "Interpretation of the Kurtosis Statistic". In: *The American Statistician* 24.4 (1970), pp. 19–22. DOI: 10.1080/00031305.1970.10477202 (cit. on p. 69).

[222] L. F. Santos, F. Borgonovi, and G. L. Celardo. "Cooperative Shielding in Many-Body Systems with Long-Range Interaction". In: *Phys. Rev. Lett.* 116 (25 June 2016), p. 250402. DOI: 10.1103/PhysRevLett.116.250402 (cit. on p. 72).

[223] A. Karpathy. *Software 2.0*. Nov. 2017. URL: https://medium.com/@karpathy/software-2-0-a64152b37c35 (cit. on p. 74).

[224] F. Rosenblatt. "The Perceptron–a perceiving and recognizing automaton". In: *Cornell Aeronautical Laboratory Report* 85 (1957) (cit. on p. 74).

[225] M. Minsky and S. Papert. *Perceptrons: An Introduction to Computational Geometry*. The MIT Press Ltd, 1969. ISBN: 0262630222. URL: https://books.google.fr/books?id=PLQ5DwAAQBAJ (cit. on p. 75).

[226] S. Linnainmaa. "The representation of the cumulative rounding error of an algorithm as a Taylor expansion of the local rounding errors." MA thesis. Univ. Helsinki, 1970 (cit. on p. 75).

[227] Wikipedia. *Harward cost of FLOPS performance*. July 2019. URL: https://en.wikipedia.org/wiki/FLOPS#Hardware_costs (cit. on p. 75).

[228] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016. ISBN: 0262035618. URL: http://www.deeplearningbook.org (cit. on pp. 75, 76).

[229] A. N. Kolmogorov. "On representation of continuous functions of several variables by superpositions of continuous functions of a smaller number of variables". In: *Doklady Akademii Nauk SSSR* 108.2 (1956), pp. 179–182. URL: https://zbmath.org/0070.28301 (cit. on p. 76).

[230] V. Arnol'd. "On functions of three variables". In: *Doklady Akademii Nauk SSSR* 114 (1957), pp. 679–681. DOI: 10.1007/978-3-642-01742-1_6 (cit. on p. 76).

[231] G. Cybenko. "Approximation by superpositions of a sigmoidal function". In: *Mathematics of Control, Signals, and Systems* 2.4 (Dec. 1989), pp. 303–314. DOI: 10.1007/bf02551274 (cit. on p. 78).

[232] K. Hornik. "Approximation capabilities of multilayer feedforward networks". In: *Neural Networks* 4.2 (1991), pp. 251–257. DOI: 10.1016/0893-6080(91)90009-t (cit. on p. 78).

[233] S. Sonoda and N. Murata. "Neural network with unbounded activation functions is universal approximator". In: *Applied and Computational Harmonic Analysis* 43.2 (Sept. 2017), pp. 233–268. DOI: 10.1016/j.acha.2015.12.005 (cit. on p. 78).

[234] Z. Lu et al. "The Expressive Power of Neural Networks: A View from the Width". In: *Advances in Neural Information Processing Systems 30*. Ed. by I. Guyon et al. Curran Associates, Inc., 2017, pp. 6231–6239. URL: http://papers.nips.cc/paper/7203-the-expressive-power-of-neural-networks-a-view-from-the-width.pdf (cit. on p. 78).

[235] R. Eldan and O. Shamir. "The Power of Depth for Feedforward Neural Networks". In: (2015). arXiv: 1512.03965 [cs.LG] (cit. on p. 78).

[236] Y. S. Bengio and Y. Le Cun. "Large Scale Kernel Machines". In: MIT Press, 2007. Chap. Scaling learning algorithms towards AI (cit. on p. 78).

[237] W. Krauth, M. Caffarel, and J.-P. Bouchaud. "Gutzwiller wave function for a model of strongly interacting bosons". In: *Phys. Rev. B* 45 (6 Feb. 1992), pp. 3137–3140. DOI: 10.1103/PhysRevB.45.3137 (cit. on p. 82).

[238] R. Orús. "A practical introduction to tensor networks: Matrix product states and projected entangled pair states". In: *Annals of Physics* 349 (2014), pp. 117–158. ISSN: 0003-4916. DOI: https://doi.org/10.1016/j.aop.2014.06.013 (cit. on pp. 82, 103).

[239] F. Verstraete, V. Murg, and J. Cirac. "Matrix product states, projected entangled pair states, and variational renormalization group methods for quantum spin systems". In: *Advances in Physics* 57.2 (2008), pp. 143–224. DOI: 10.1080/14789940801912366 (cit. on p. 82).

[240] F. Mezzacapo et al. "Ground-state properties of quantum many-body systems: entangled-plaquette states and variational Monte Carlo". In: *New Journal of Physics* 11.8 (Aug. 2009), p. 083026. DOI: 10.1088/1367-2630/11/8/083026 (cit. on p. 82).

[241] W. M. C. Foulkes et al. "Quantum Monte Carlo simulations of solids". In: *Rev. Mod. Phys.* 73 (1 Jan. 2001), pp. 33–83. DOI: 10.1103/RevModPhys.73.33 (cit. on p. 82).

[242] F. Becca and S. Sorella. *Quantum Monte Carlo Approaches for Correlated Systems*. Cambridge University Press, 2017. DOI: 10.1017/9781316417041 (cit. on pp. 82, 83).

[243] F. Becca and S. Sorella. "Optimization of Variational Wave Functions". In: *Quantum Monte Carlo Approaches Correl. Syst.* Cambridge University Press, 2017, pp. 131–155. ISBN: 9781316417041. DOI: 10.1017/9781316417041.007 (cit. on pp. 82, 94).

[244] T. Kim and T. Adali. "Universal approximation of fully complex feed-forward neural networks". In: *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*. Vol. 1. May 2002, pp. I-973-I-976. DOI: 10.1109/ICASSP.2002.5743956 (cit. on p. 82).

[245] H. Saito. "Solving the Bose–Hubbard Model with Machine Learning". In: *Journal of the Physical Society of Japan* 86.9 (2017), p. 093001. DOI: 10.7566/JPSJ.86.093001 (cit. on p. 83).

[246] H. Saito and M. Kato. "Machine Learning Technique to Find Quantum Many-Body Ground States of Bosons on a Lattice". In: *Journal of the Physical Society of Japan* 87.1 (Jan. 2018), p. 014001. DOI: 10.7566/jpsj.87.014001 (cit. on p. 83).

[247] A. Bansal et al. "PixelNet: Representation of the pixels, by the pixels, and for the pixels". In: (2017). arXiv: 1702.06506 [cs.CV] (cit. on p. 83).

[248] O. Sharir et al. "Deep autoregressive models for the efficient variational simulation of many-body quantum systems". In: (2019). arXiv: 1902.04057 [cond-mat.dis-nn] (cit. on pp. 83, 90, 91, 101).

[249] G. Torlai and R. G. Melko. "Latent Space Purification via Neural Density Operators". In: *Phys. Rev. Lett.* 120 (24 June 2018), p. 240503. DOI: 10.1103/PhysRevLett.120.240503 (cit. on pp. 83, 84).

[250] N. L. Roux and Y. Bengio. "Representational Power of Restricted Boltzmann Machines and Deep Belief Networks". In: *Neural Computation* 20.6 (June 2008), pp. 1631–1649. DOI: 10.1162/neco.2008.04-07-510 (cit. on p. 85).

[251] L. Younes. "Synchronous Boltzmann machines can be universal approximators". In: *Applied Mathematics Letters* 9.3 (May 1996), pp. 109–113. DOI: 10.1016/0893-9659(96)00041-9 (cit. on p. 85).

[252] G. Montufar and N. Ay. "Refinements of Universal Approximation Results for Deep Belief Networks and Restricted Boltzmann Machines". In: *Neural Computation* 23.5 (May 2011), pp. 1306–1319. DOI: 10.1162/neco_a_00113 (cit. on p. 85).

[253] R. Bhatia. *Matrix Analysis*. Ed. by Springer-Verlag. Vol. 169. New York: Springer New York, 1997. DOI: 10.1007/978-1-4612-0653-8 (cit. on p. 88).

[254] S. Kumar, M. Mohri, and A. Talkwalkar. "On Sampling-Based Approximate Spectral Decomposition". In: *International Conference on Machine Learning (ICML)*. 2009. URL: http://www.sanjivk.com/nys_col_ICML.pdf (cit. on p. 88).

[255] P.-G. Martinsson. "Randomized methods for matrix computations". In: (2015). arXiv: 1607.01649 [math.NA] (cit. on p. 88).

[256] P. Drineas and M. W. Mahoney. "On the Nyström method for approximating a Gram matrix for improved kernel-based learning". In: *journal of machine learning research* 6.Dec (2005), pp. 2153–2175. URL: http://www.jmlr.org/papers/volume6/drineas05a/drineas05a.pdf (cit. on p. 88).

[257] C. J. Geyer. "Practical Markov Chain Monte Carlo". In: *Statistical Science* 7.4 (Nov. 1992), pp. 473–483. DOI: 10.1214/ss/1177011137 (cit. on p. 90).

[258] W. Gilks, S. Richardson, and D. Spiegelhalter. *Markov Chain Monte Carlo in Practice*. Chapman & Hall/CRC Interdisciplinary Statistics. Taylor & Francis, 1995. ISBN: 9780412055515. URL: https://books.google.fr/books?id=TRXrMWY%5C_i2IC (cit. on p. 90).

[259] M. Norouzi, M. Ranjbar, and G. Mori. "Stacks of convolutional restricted boltzmann machines for shift-invariant feature learning". In: *2009 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work. CVPR Work. 2009*. Vol. 2009 IEEE. IEEE, June 2009, pp. 2735–2742. ISBN: 9781424439935. DOI: 10.1109/CVPRW.2009.5206577. arXiv: 1203.4416 (cit. on p. 91).

[260] K. Sohn and H. Lee. "Learning Invariant Representations with Local Transformations". In: *Proc. 29th Int. Coference Int. Conf. Mach. Learn.* Omnipress, 2012, pp. 1339–1346. ISBN: 978-1-4503-1285-1. DOI: 1206.6418. arXiv: 1206.6418 (cit. on p. 91).

[261] T. Vieijra et al. "Restricted Boltzmann Machines for Quantum States with Non-abelian or Anyonic Symmetries". In: (2019). arXiv: 1905.06034 [cond-mat.str-el] (cit. on p. 91).

[262] C. J. Umrigar and C. Filippi. "Energy and Variance Optimization of Many-Body Wave Functions". In: *Phys. Rev. Lett.* 94 (15 Apr. 2005), p. 150201. DOI: 10.1103/PhysRevLett.94.150201 (cit. on p. 92).

[263] A. R. Conn, K. Scheinberg, and L. N. Vicente. *Introduction to Derivative-Free Optimization (Mps-siam Series on Optimization)*. Society for Industrial and Applied Mathematics, 2009. ISBN: 0898716683 (cit. on p. 92).

[264] D. Karaboga and B. Basturk. "A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm". In: *Journal of global optimization* 39.3 (2007), pp. 459–471 (cit. on p. 92).

[265] H. Szu and R. Hartley. "Fast simulated annealing". In: *Physics Letters A* 122.3-4 (June 1987), pp. 157–162. DOI: 10.1016/0375-9601(87)90796-1 (cit. on p. 92).

[266] L. Bottou. "Stochastic Gradient Descent Tricks". In: *Neural Networks: Tricks of the Trade: Second Edition*. Ed. by G. Montavon, G. B. Orr, and K.-R. Müller. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 421–436. ISBN: 978-3-642-35289-8. DOI: 10.1007/978-3-642-35289-8_25 (cit. on p. 93).

[267] Y. Dauphin et al. "Identifying and attacking the saddle point problem in high-dimensional non-convex optimization". In: (2014). arXiv: 1406.2572 [cs.LG] (cit. on p. 93).

[268] P. Chaudhari and S. Soatto. "On the energy landscape of deep networks". In: (2015). arXiv: 1511.06485 [cs.LG] (cit. on p. 93).

[269] I. Sutskever et al. "On the importance of initialization and momentum in deep learning". In: *International conference on machine learning*. 2013, pp. 1139–1147 (cit. on p. 93).

[270] R. Pascanu and Y. Bengio. "Revisiting Natural Gradient for Deep Networks". In: (2013). arXiv: 1301.3584 [cs.LG] (cit. on p. 94).

[271] J. Martens. "New insights and perspectives on the natural gradient method". In: (2014). arXiv: 1412.1193 [cs.LG] (cit. on p. 94).

[272] S.-C. T. Choi and M. A. Saunders. "Algorithm 937". In: *ACM Transactions on Mathematical Software* 40.2 (Mar. 2014), pp. 1–12. DOI: 10.1145/2527267 (cit. on p. 95).

[273] F. Vicentini. *GitHub reporsitory: NeuralQuantum.jl*. 2019. URL: https://github.com/PhilipVinc/NeuralQuantum.jl (cit. on p. 96).

[274] F. Vicentini. *PhilipVinc/NeuralQuantum.jl*. 2019. DOI: 10.5281/ZENODO.3238294 (cit. on pp. 96, 115).

[275] J. Jin et al. "Phase diagram of the dissipative quantum Ising model on a square lattice". In: *Phys. Rev. B* 98 (24 Dec. 2018), p. 241108. DOI: 10.1103/PhysRevB.98.241108 (cit. on p. 98).

[276] H. Li et al. "Visualizing the Loss Landscape of Neural Nets". In: (2017). arXiv: 1712.09913 [cs.LG] (cit. on pp. 98, 99).

[277] F. Arute et al. "Quantum supremacy using a programmable superconducting processor". In: *Nature* 574.7779 (Oct. 2019), pp. 505–510. DOI: 10.1038/s41586-019-1666-5 (cit. on p. 105).

[278] M. Innes. "Don't Unroll Adjoint: Differentiating SSA-Form Programs". In: (2018). arXiv: 1810.07951 [cs.LG] (cit. on p. 113).

[279] G. Guennebaud, B. Jacob, et al. *Eigen v3*. 2010. URL: http://eigen.tuxfamily.org (cit. on p. 115).

[280] C. C. Margossian. "A review of automatic differentiation and its efficient implementation". In: *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* (Mar. 2019), e1305. DOI: 10.1002/widm.1305 (cit. on p. 115).