Periodically-driven quantum many-body systems, many-body localization and machine learning

by

Pedro Ponte

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Physics

Waterloo, Ontario, Canada, 2018

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

Supervisor	**Roger Melko** Associate Professor, Department of Physics and Astronomy, University of Waterloo Associate Faculty, Perimeter Institute for Theoretical Physics
External Examiner	**Richard Scalettar** Professor, University of California, Davis
Internal Member	**Jeff Z. Y. Chen** Professor, Department of Physics and Astronomy, University of Waterloo
Internal-external Member	**Mohammad Kohandel** Associate Professor, Department of Applied Mathematics, University of Waterloo
Committee Member	**Anton Burkov** Associate Professor, Department of Physics and Astronomy, University of Waterloo

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of contributions

Chapter 2 of this thesis consists of material from Reference [1], co-authored with Anushya Chandran, Zlatko Papić, and Dmitry A. Abanin.

Chapter 3 consists of material from Reference [2], co-authored with Zlatko Papić, François Huveneers, and Dmitry A. Abanin.

Chapter 4 consists of material from Reference [3], co-authored with Christopher R. Laumann, David A. Huse, and Anushya Chandran.

Chapter 7 consists of material from Reference [4], co-authored with Roger G. Melko.

Abstract

Despite the exponentially large amount of information required in the quantum description of many-body systems, finite size numerical simulations have paved the way for recent progress in the understanding of quantum phases of matter both in and out-of-equilibrium.

In this spirit, using exact diagonalization for a small number of spins ~ 20, first, we investigate the fate of both ergodic quantum many-body systems that thermalize at long times and systems with quenched disorder that exhibit many-body localization (MBL) and fail to thermalize (due to an extensive set of local conserved degrees of freedom) under periodic driving. Ergodic systems always delocalize in energy space and heat up to infinite temperature, for both local driving, in which the time dependence of the Hamiltonian is restricted to a few contiguous degrees of freedom, and global driving. On the other hand, MBL systems remain localized at finite energy density for local driving, while in the case of global driving these systems can either heat up to infinite temperature or remain localized depending on the parameters of the drive. Underlying the latter is the emergence of an effective time independent MBL Hamiltonian describing the long time dynamics of the system. These numerical predictions have recently been verified experimentally in a one dimensional quasi-disordered optical lattice of interacting fermions wherein unitary time evolution can be experimentally probed.

Subsequently, we study the stability of MBL systems with respect to another perturbation, in this case by thermal inclusions, which remains poorly understood. We consider a simple model to address this question: a two level system interacting with N localized bits subject to random fields. In the large N limit, the localized bits are thermalized for any nonzero ratio of interactions to disorder strength. Moreover, at finite N, in the transition region, the single-site eigenstate entanglement entropies exhibit bimodal distributions, so that localized bits are either "on" (strongly entangled) or "off" (weakly entangled) in eigenstates. The clusters of "on" bits vary significantly between eigenstates of the same sample, which provides evidence for a heterogenous discontinuous transition out of the localized phase in single site observables. These results also imply the MBL phase is unstable in systems with short-range interactions and quenched randomness in dimensions d that are high but finite.

The second part of this thesis is motivated by the recent breakthroughs of artificial intelligence in domains such as computer vision involving highly dimensional data of complexity not unlike the one found in the description of many-body systems. Using Monte Carlo simulations to generate synthetic data, we find that kernel methods, a class of machine learning algorithms, can learn the mathematical form of physical discriminators associated

with phase transitions, such as order parameters and Hamiltonian constraints, for a set of two dimensional spin models: the ferromagnetic Ising model, a conserved-order-parameter Ising model, and the Ising gauge theory.

Our work shows that the combination of numerical simulations with the ability of machine learning algorithms for automatic feature detection hold promise to become standard tools in the exploration and characterization of different phases of matter in condensed matter and other many-body systems.

Acknowledgements

First and foremost I thank my supervisor Roger Melko for his supervision and support during these years. I look up to you not only as an elite academic, but also for being a great mentor who cares for and empowers his students.

I am grateful to have as collaborators an incredible set of physicists: Dmitry Abanin, Anushya Chandran, David Huse, Chris Laumann and Zlatko Papic. I would like to acknowledge all the interesting ideas, discussions and guidance that I got from our interactions. In particular, I thank Anushya and Chris for the superb hospitality during the time I spent in Boston and also their mentorship during this time. I am also grateful to Anton Burkov, Sung-Sik Lee, Ali Ghodsi, and Brian McNamara for serving on my committee meetings and providing valuable feedback. Moreover, a word of appreciation to Judy McDonnell and Debbie Guenther is more than deserved for always being available to help me tackle the bureaucratic demands of the program.

I also thank all my close friends and family for the support you provided. In particular, I deeply appreciate and I will always be indebted to my parents and brother for their unconditional love and patience during this challenging journey.

Finally, a crucial thank you to Paulina. You were the linchpin of my support. Thank you for constantly pushing me to appreciate and do more of what is truly meaningful in life (despite the often lack of success), and for all the love, joy, happiness and serendipity that you embedded into my life.

Table of Contents

List of Figures

Part I

Instabilities in ergodic and many-body localized systems

Chapter 1

Introduction

1.1 Thermalization

1.1.1 Classical systems

Statistical Mechanics studies the laws governing macroscopic bodies formed by a formidable number of individual particles $N \gg O(10^{23})$. From a classical viewpoint, each particle is described by its position and velocity (or momentum) at each instant of time. The set of positions and momenta of the particles $\boldsymbol{x}_i, \boldsymbol{p}_i$ defines the so-called phase space of the system. The dynamics of the system, i.e. how it evolves through time, is determined by the Hamiltonian that describes at the microscopic level how the particles interact and how they move as a consequence, following the laws of Newtonian mechanics. Thus, if a complete microscopic description is available, equations of motion for each of the particles can be readily obtained. These can be solved and provide complete information about the state of the system at any future instant in time.

Nevertheless, one immediately understands the challenge of such attempt. Even in the uncommon situation where an analytical solution is available, just the first step of holding in computer memory the initial state for every particle is far from reach for current technology. Moreover, with the natural expectation that the complexity in the behavior of macroscopic bodies only increases with the number of particles, any progress on the problem of deriving the macroscopic behaviour from the microscopic dynamics is apparently doomed to failure.

Yet, notwithstanding that large systems are described by the same microscopic laws as systems composed of only a small number of particles, the large number of interacting

degrees of freedom motivate a powerful hypothesis that simplifies the problem and is able to explain a large range of macroscopic phenomena.

The ergodic hypothesis [6] states that due to the complex microscopic interactions, during a sufficiently long time, the system visits every possible state compatible with macroscopic conservation laws, e.g. energy. Additionally, the time spent in a particular state is proportional to its volume element in phase space. Mathematically, this means the probability of the system being in the state defined by the set of positions and momenta $\boldsymbol{x}_1, \boldsymbol{p}_1, \ldots, \boldsymbol{x}_N, \boldsymbol{p}_N$ is proportional to a delta-function: $\delta\left(E(\boldsymbol{x}_1, \boldsymbol{p}_1, \ldots, \boldsymbol{x}_N, \boldsymbol{p}_N) - E_0\right)$. This probability distribution defines the microcanonical ensemble, where E_0 is the energy associated with the initial state of the system. This ensemble is compatible with an equilibrium description as an initial set of systems sampled according to its distribution still satisfies the same distribution at later times; this statement is known as the Liouville theorem [5]. The ergodic hypothesis shifts the paradigm from a microscopic description to a macroscopic one, whereby the system is described by the few macroscopic observables that are conserved by its dynamics.

Thus, according to the ergodic hypothesis the average value of a physical observable O observed over sufficiently long times, i.e. $\bar{O} = \lim_{t\to\infty} \frac{1}{t} \int_0^t dt O(t)$, is equal to the microcanonical average:

$$\bar{O} = \frac{\int d\boldsymbol{x}_1 d\boldsymbol{p}_1 \ldots d\boldsymbol{x}_N d\boldsymbol{p}_N O(\boldsymbol{x}_1, \boldsymbol{p}_1, \ldots, \boldsymbol{x}_N, \boldsymbol{p}_N) \delta\left(E(\boldsymbol{x}_1, \boldsymbol{p}_1, \ldots, \boldsymbol{x}_N, \boldsymbol{p}_N) - E_0\right)}{\int d\boldsymbol{x}_1 d\boldsymbol{p}_1 \ldots d\boldsymbol{x}_N d\boldsymbol{p}_N \delta\left(E(\boldsymbol{x}_1, \boldsymbol{p}_1, \ldots, \boldsymbol{x}_N, \boldsymbol{p}_N) - E_0\right)} \tag{1.1}$$

Moreover, the average value $\bar{O}$ provides a very accurate description of the equilibrium state for additive physical quantities, as, in the thermodynamic limit, deviations from this value are *very* unlikely [5]. This result relies on the statistical independence of $O(N)$ macroscopic subsystems in which the full system can be divided, which are still large enough to behave as macroscopic bodies. This assumption is valid as long as interactions are approximately local and for short enough periods of time, before interaction effects ultimately propagate into the subsystem. Under the assumption of statistical independence, the ratio of the time fluctuations of the physical observable O to its mean value is suppressed in the thermodynamic limit as $1/\sqrt{N}$. Consequently, most configurations with significant weight in the microcanonical ensemble bear similar values with respect to macroscopic observables.

The ergodic hypothesis provides the connection between microscopic and macroscopic behavior on which the field of Thermodynamics relies to explain an incredible range of phenomena. For example, a hot cup of coffee eventually cools down because the number of microscopic configurations compatible with the temperature of the coffee and its environment being the same is overwhelmingly larger than in any other available scenario. This

is an example of the second law of thermodynamics at play, which states that for a closed system in a non-equilibrium state the most likely sequence of events is a steady increase of the system's entropy.

However, it should be acknowledged that there is still debate in the community about the finer details connecting microscopic to macroscopic thermodynamic behavior in classical systems [6].

1.1.2 Quantum systems

Numerous independent experiments have shown that an accurate description of reality must take into account quantum effects. Consequently, the relationship between macroscopic thermalization phenomena and the microscopic quantum mechanical description must be understood. As we shall see in the following, this relationship between the microscopic dynamics and the emergence of a statistical mechanical description is more subtle in quantum systems and qualitatively different from the classical case.

In this thesis, we will focus on isolated quantum systems. A complete description of these systems requires the specification of the Hilbert Space defining the possible states of the system and the Hamiltonian $H(t)$ which determines the unitary time evolution of the system according to the Schrodinger equation:

$$i\frac{d|\psi(t)\rangle}{dt} = H|\psi(t)\rangle. \tag{1.2}$$

Among the most common protocols for driving quantum systems out of equilibrium are the so-called quenches. This means that the system is initially prepared in a state that is either simple to prepare (e.g. product state) or corresponds to an equilibrium state or ground state of another Hamiltonian H'. Quenches in which H' differs from H in parameters throughout the system are called global quenches, while quenches in which parameters are only changed in a finite region are called local quenches. The subsequent discussion considers initial states $|\varphi_0\rangle$ that are pure for simplicity, however the generalization to initial mixed states is trivial. The initial states $|\varphi_0\rangle$ in quench experiments typically have a well-defined energy density $\varepsilon_0 = \langle\varphi_0|H|\varphi_0\rangle/N$, where N is the number of degrees of freedom, in the sense that the energy fluctuations $\Delta E_0 = \sqrt{\langle\varphi_0|H^2|\varphi_0\rangle - \langle\varphi_0|H|\varphi_0\rangle^2}$ are sub-extensive [7]. This is an important requirement, so that the long time limit behaviour can be compared to the thermodynamic ensembles where this property holds.

The time evolution of the initial state $|\varphi_0\rangle$ when expanded in the basis of eigenstates

of the Hamiltonian $|E_m\rangle$ is simply:

$$|\psi(t)\rangle = \sum_m A_m e^{-iE_m t}|E_m\rangle, \tag{1.3}$$

where E_m is the eigenenergy of the Hamiltonian associated with eigenstate $|E_m\rangle$ and $A_m = \langle E_m|\varphi_0\rangle$ denotes the projection of the initial state on the eigenstate basis.

As the framework of statistical mechanics relies on an ensemble description, it is natural to move from a wavefunction to a density matrix description of the system. In this formulation, the state of the system is

$$\rho(t) = |\psi(t)\rangle\langle\psi(t)| = \sum_m |A_m|^2|E_m\rangle\langle E_m| + \sum_{m\neq n} A_m A_n^* e^{-i(E_m-E_n)t}|E_m\rangle\langle E_n|, \tag{1.4}$$

and the time evolution of an observable $\mathcal{O}$ is given by

$$\mathcal{O}(t) = \mathrm{Tr}\rho(t)\mathcal{O} = \sum_m |A_m|^2\mathcal{O}_{mm} + \sum_{m\neq n} A_m A_n^* \mathcal{O}_{nm} e^{-i(E_m-E_n)t}, \tag{1.5}$$

where $\mathcal{O}_{mn} = \langle E_m|\mathcal{O}|E_n\rangle$. From Eq. 1.5, we can enumerate the conditions under which generic observables thermalize in the long time limit. For an observable to thermalize, the equilibrium value $\bar{\mathcal{O}}$ given by the time average $\bar{\mathcal{O}} = \lim_{T\to\infty}\frac{1}{T}\int_0^T dt\mathcal{O}(t) = \sum_m |A_m|^2\mathcal{O}_{mm}$ must agree with the predictions of the microcanonical ensemble ρ_{MC} and the canonical ensemble ρ_{C}, which we define shortly. Notice that $\bar{\mathcal{O}}$ is the expectation value in the so-called diagonal ensemble:

$$\rho_D = \sum_m |A_m|^2|E_m\rangle\langle E_m|. \tag{1.6}$$

Further, for thermalization to occur in the strong sense [6] the expectation value $\mathcal{O}(t)$ must remain close to the equilibrium value $\bar{\mathcal{O}}$ for most of the time, which requires the time dependent contributions in Eq. 1.5 to vanish at long times in the thermodynamic limit.

Here it is assumed that energy is the only extensive conserved quantity (however, other possible conservation laws might exist, e.g. particle number, total spin, etc). The microcanonical ensemble is the maximum entropy density matrix restricted to a macroscopically small energy shell $[E_0 - \Delta E_0/2, E_0 + \Delta E_0/2]$ with, e.g. $\Delta E = O(N^0)$, around the mean energy of the initial state $E_0 = \langle\varphi_0|H|\varphi_0\rangle$:

$$\rho_{\mathrm{MC}} = \frac{1}{\mathcal{N}} \sum_{m:E_m\in[E_0-\Delta E/2, E_0+\Delta E/2]} |E_m\rangle\langle E_m|, \tag{1.7}$$

where $\mathcal{N}$ is the number of eigenstates in the energy shell.

The thermal/canonical ensemble is given by the usual Gibbs distribution

$$\rho_{\mathrm{C}} = \frac{1}{Z_\beta} e^{-\beta H}, \tag{1.8}$$

where the partition function is defined as $Z_\beta = \mathrm{Tr} e^{-\beta H}$, and the inverse temperature $\beta = 1/T$ is fixed by the mean energy of the initial state $E_0 = \langle \varphi_0 | H | \varphi_0 \rangle$ such that $E_0 = \mathrm{Tr} \rho_{\mathrm{C}} H$ is verified.

A natural definition of thermalization would be the requirement that the density matrices $\rho_D, \rho_{\mathrm{MC}}, \rho_{\mathrm{C}}$ converge in the thermodynamic limit:

$$\rho_D = \rho_{\mathrm{MC}} = \rho_{\mathrm{C}}. \tag{1.9}$$

Consequently, if Eq. 1.9 holds, all observables converge to their thermal values at long times. However, this is not the case for generic initial states. As we will see, thermalization in quantum systems is understood through the behaviour of local observables as opposed to considering the state of the entire system.

Eigenstate Thermalization Hypothesis

The Eigenstate Thermalization Hypothesis [9, 10, 11, 7] defines an ansatz for the matrix elements of local observables $\mathcal{O}$ in the basis of the eigenstates $|E_m\rangle$ of the Hamiltonian:

$$\mathcal{O}_{mn} = \mathcal{O}(\bar{E}) \delta_{mn} + e^{-S(\bar{E})/2} f_{\mathcal{O}}(\bar{E}, \omega) R_{nm}. \tag{1.10}$$

Here $\bar{E} = (E_n + E_m)/2$, $\omega = E_n - E_m$, and $S(\bar{E}) \sim \ln(\rho(\bar{E}))$ is the thermodynamic entropy at energy $\bar{E}$ defined in terms of the density of states ρ that counts the number of eigenstates per energy interval. Moreover, both $\mathcal{O}(\bar{E})$ and $f_{\mathcal{O}}(\bar{E}, \omega)$ are smooth functions of their arguments and R_{nm} is a random variable with mean zero and variance $O(1)$. By now, numerical simulations have verified the ETH on several distinct one dimensional systems [12, 13, 14, 15, 6]. Quantum many-body systems that satisfy ETH are also known as ergodic systems.

The ETH ansatz is motivated by Random Matrix Theory (RMT) [6]. Eigenstates of Hamiltonians sampled from the RMT ensembles are uncorrelated random states and, under this assumption, the matrix elements of a local observable satisfy $\mathcal{O}_{mn} = \langle O \rangle \delta_{mn} + \sqrt{\frac{\langle \mathcal{O}^2 \rangle}{\mathcal{D}}} R_{mn}$ [6]; $\mathcal{D}$ is the dimension of the Hilbert space and the averages $\langle \mathcal{O} \rangle$, $\langle \mathcal{O}^2 \rangle$ are basis independent, e.g. $\langle \mathcal{O} \rangle = \frac{1}{\mathcal{D}} \sum_{n=1}^{\mathcal{D}} \mathcal{O}_{nn}$. ETH adapts RMT by arguing that eigenstates of

local Hamiltonians, which by definition are characterized by a given energy, are essentially random after taking this constraint into account. This means that the diagonal matrix element $\mathcal{O}_{nn}$ is allowed to be a function of the energy, and the typical value of the off-diagonal elements $\mathcal{O}_{mn}$ is set by the available number of eigenstates with energy $\sim \bar{E}$, which is proportional to the density of states $\rho(\bar{E}) \sim e^{S(\bar{E})}$, instead of the total number of states in the Hilbert Space. These arguments suggest the form of Eq. 1.10. Moreover, the function $f_{\mathcal{O}}(\bar{E}, \omega)$ must decay rapidly with ω on the scale of the energy per degree of freedom as a local operator only affects a few degrees of freedom and can only change the energy by a finite amount.

Using the ETH ansatz, we can obtain an expression for the equilibrium value $\bar{\mathcal{O}}$ [6]:

$$\bar{\mathcal{O}} = \sum_m |A_m|^2 \mathcal{O}_{mm} \tag{1.11}$$

$$= \sum_m |A_m|^2 \mathcal{O}(E_m) + \dots \tag{1.12}$$

$$\approx \sum_m |A_m|^2 \mathcal{O}(E_0 + \delta E_m) \tag{1.13}$$

$$\approx \sum_m |A_m|^2 \left[\mathcal{O}(E_0) + \frac{d\mathcal{O}}{dE} \delta E_m + \frac{1}{2} \frac{d^2\mathcal{O}}{dE^2} (\delta E_m)^2 + \dots \right] \tag{1.14}$$

$$\approx \mathcal{O}(E_0) + \frac{1}{2} \frac{d^2\mathcal{O}}{dE^2} (\Delta E_0)^2 \tag{1.15}$$

In the second line, the random terms in the diagonal matrix elements were neglected, since they account for an exponentially small contribution. And, in the fourth line, $\mathcal{O}(E_0 + \delta E_m)$ from Eq. 1.10 is expanded up to second order terms in δE_m.

An analogous calculation shows the expectation values $\mathcal{O}_{\text{MC}} = \text{Tr}\rho_{\text{MC}}\mathcal{O}$, $\mathcal{O}_{\text{C}} = \text{Tr}\rho_{\text{C}}\mathcal{O}$ differ from $\bar{\mathcal{O}}$ only in the second term in Eq. 1.15 due to each ensemble having different energy uncertainties ΔE. Moreover, since the energy uncertainties are sub-extensive in any of the ensembles and for local operators the function $\mathcal{O}(E_m)$ only depends explicitly on the energy density [16], the second term vanishes with increasing system size. Thus, in the thermodynamic limit, the three ensembles are in perfect agreement: $\bar{\mathcal{O}} = \mathcal{O}_{\text{MC}} = \mathcal{O}_{\text{C}}$.

Moreover, the ETH predicts that time fluctuations of an observable $\Delta\mathcal{O}^2 = \lim_{t\to\infty} \frac{1}{t} \int_0^t dt \left(\mathcal{O}(t) - \bar{\mathcal{O}}\right)^2$ are exponentially small in the number of degrees of freedom (particles, spins, ...) [6]. This implies that at long times observables remain near their thermal values most of the time, thus thermalization occurs in the strong sense.

However, ETH is not valid for certain systems that also fail to thermalize. These can be divided into two broad categories, namely integrable systems and many-body localization.

The latter is the only robust example though, as it remains non-thermalizing when generic small perturbations are added to the Hamiltonian, i.e. it defines a phase.

Integrable systems

Integrable systems include all non-interacting models, interacting models that can be mapped to non-interacting ones, and specific interacting models that cannot be mapped to non-interacting ones.

For concreteness, an example of the first category is a free fermion system with nearest neighbour hopping on a one dimensional chain: $H = -t \sum_{i=1}^{N} \left(c_{i+1}^{\dagger} c_i + h.c. \right)$. Any generic non-interacting Hamiltonian such as this one can be expressed in the diagonal form:

$$H = \sum_{m=1}^{N} E_m I_m, \tag{1.16}$$

where the particle numbers $I_m = c_m^{\dagger} c_m$, are defined by the creation/annihilation operators c_m constructed from the single-particle states that diagonalize the Hamiltonian and the E_m are the associated eigenenergies; for this particular example, the c_m correspond to momentum eigenstates.

The 1d Transverse Field Ising Model constitutes an example for the second category, since it can be cast in the same diagonal form as Eq. 1.16, after a Jordan-Wigner transformation mapping the 1/2-spin degrees of freedom to fermions followed by a Bogoliubov transformation [17].

The $O(N)$ quantities $I_m = c_m^{\dagger} c_m$ commute with the Hamiltonian and with each other, and thus are independently conserved by the dynamics after a quench experiment. It was proposed and verified in numerical simulations [18] that these systems equilibrate at long times to the predictions of the Generalized Gibbs Ensemble $\rho_{\mathrm{GGE}} \sim e^{-\sum_{m=1}^{N} \ell_m I_m}$. This ensemble maximizes the entropy under the constrain that it reproduces the expected values of the conserved quantities I_n in the initial state. The only reported case where the GGE failed was in the presence of disorder [19].

At first sight, there are a few important criticisms to the GGE [8]. The I_m can be non-local, for example when these correspond to the occupation of momentum eigenstates, which is the case both for the free fermion model and the TFIM mentioned above. This poses a problem, since including non-local conserved quantities in the ensemble, means it would be admissible to include the exponential number of eigenstate projectors in a thermal ensemble for generic quantum many-body systems. Moreover, it is very important to deal

with local additive conserved quantities like the energy, so that the statistical mechanics assumption of approximate subsystem independence applies [5]. However, it turns out the non-local I_m can be expressed in terms of local additive quantities, which settles the issue [6]. This also hints at a generalization of the ETH [20, 6], whereby the values of local observables are smooth functions of not just the energy, but also of the other local extensive conserved quantities.

1.2 Many-body localization

Many-body localization [21, 22] has its origins in single-particle systems. When the Hamiltonian contains some *randomness*, non-interacting systems can exhibit interesting localization phenomena known as Anderson localization [29].

For concreteness, consider a one dimensional chain of fermions with nearest neighbour hopping:

$$H = -t \sum_i (c_{i+1}^\dagger c_i + h.c.) + \sum_i W_i c_i^\dagger c_i, \tag{1.17}$$

where the disorder terms W_i are sampled form the uniform distribution on the interval $[-W, W]$ defined by the disorder strength W.

Through diagonalization, H can be cast in a diagonal form: $H = \sum_m E_m c_m^\dagger c_m$. The single-particle eigenstates $|E_m\rangle$ are localized in real space centered around an eigenstate dependent location r_m. The amplitude for finding the fermion a distance R away from r_m decays exponentially on the scale of the localization length ξ, i.e. it is proportional to $e^{-R/\xi}$.

In the limit of zero hopping, the eigenstates are strictly localized on each lattice size. In the perturbative regime $t/W \ll 1$ of small hopping, localization survives, as from first order perturbation theory, a given site only has a small probability of hybridizing with neighbouring sites. Turns out, in dimensions one and two, all eigenstates are localized in the thermodynamic limit for arbitrarily small disorder strength W. In higher dimensions, there is a transition between localized and delocalized eigenstates. This can happen as a function of disorder strength or energy density. In the latter, localized states at the edge of the spectrum are separated by a mobility edge from delocalized states, and these critical states show a power-law localization profile.

The localization of even some of the single-particle states in the spectrum has consequences for thermalization. Consider a quench in which the initial state contains an excess

density of particles in a region of localized eigenstates. Due to the conservation of the number of particles in each of the localized states, the particles in the localized region do not diffuse *completely*, so the system fails to thermalize [21].

Thus, disorder through localization phenomena prevents thermalization in single-particle systems. It is then natural to ask to what extent localization survives when weak interactions are introduced in the system.

Recent works based on perturbation theory showed that localization is indeed stable to weak, short-ranged interactions [23, 24]. Subsequently, many numerical studies of one dimensional models for 1/2-spins and fermions further supported these findings.

To clarify the meaning of localization in the many-body context, it is useful to focus on a concrete model. Most numerical studies have adopted the following 1/2-spin model:

$$H = J\sum_{i=1}^{N-1}\left(S_i^x S_{i+1}^x + S_i^y S_{i+1}^y\right) + J\sum_{i=1}^{N-1} S_i^z S_{i+1}^z + \sum_{i=1}^{N} h_i S_i^z, \tag{1.18}$$

where the spin operators are defined in terms of the Pauli matrices σ_i^α as $S_i^\alpha = \sigma_i^\alpha/2$ with $\alpha = x, y, z$, J is the hopping/interaction strength, and h_i is sampled from the uniform distribution $[-W, W]$ representing the disorder in the system. Through a Jordan-Wigner transformation this model represents interacting fermions moving on a disordered potential, and differs from Eq. 1.17 due to nearest neighbour interactions.

In the limit $J = 0$, the many-body eigenstates have a simple product state form $|s_1 s_2 \dots s_n\rangle \equiv |s_1\rangle \otimes |s_2\rangle \otimes \dots \otimes |s_N\rangle$, where each spin is aligned along the z direction with two possible orientations $s_i = \langle\sigma_i^z\rangle = \pm 1$. Moreover, the $O(N)$ operators σ_i^z form a set of conserved quantities, which commute with the Hamiltonian and with each other. At larger disorder, but non-zero hopping and interactions $J/W \ll 1$, Eq. 1.18 remains in the fully many-body localized regime. All eigenstates of the Hamiltonian are localized as each spin only hybridizes weakly with its neighbours, even though the set of operators σ_i^z no longer commute with the Hamiltonian. In fact, many-body localization is robust for a larger parameter range until $J \approx W$ [30, 31].

It was argued [25, 26] that at strong enough disorder a quasi-local unitary operator diagonalizes the Hamiltonian in the product state basis $|s_1 s_2 \dots s_N\rangle$ associated with σ_i^z: $H_{\text{diag}} = U^\dagger H U$. The quasi-locality of U means that it cannot modify in a correlated way degrees of freedom which are far away on the scale of a many-body localization length ξ. For example, when U is applied to a product state $|s_1 s_2 \dots s_N\rangle$, the overlap with states that differ from the original one by spin-flips involving spins separated by a distance R decays exponentially with R on the scale ξ.

Consequently, we can define localized two-state degrees of freedom τ_i^z called l-bits (i.e. localized-bits) [26] related to the physical spins σ_i^z by $\tau_i^z = U\sigma_i^z U^\dagger$. By definition, the τ_i^z commute with the original Hamiltonian H and with each other. The quasi-locality of U implies the following real space profile for a given l-bit:

$$\tau_i^z = \sum_{j,\alpha} K_{ij}^{\alpha}\sigma_j^{\alpha} + \sum_{jk,\alpha\beta} K_{ijk}^{\alpha\beta}\sigma_j^{\alpha}\sigma_k^{\beta} + \ldots, \tag{1.19}$$

where the coefficients K_{ij}^{α}, $K_{ijk}^{\alpha\beta}$ decay with the maximum distance between any pairs formed by spin i and the physical spins involved in the corresponding term. The omitted higher order terms behave in an analogous way.

Because by definition H_{diag} depends only on the σ_i^z and their products, the original Hamiltonian H has the following form in terms of the l-bits τ_i^z [26, 21]:

$$H = \sum_i J_i^{(1)}\tau_i^z + \sum_{ij} J_{ij}^{(2)}\tau_i^z\tau_j^z + \sum_{ijk} J_{ijk}^{(3)}\tau_i^z\tau_j^z\tau_k^z + \ldots \tag{1.20}$$

Again, due to the quasi-locality of U, the coefficients of multi-spin products in the Hamiltonian decay exponentially with the maximum distance between every pair of spins involved.

Physically, the form of the Hamiltonian Eq. 1.20 is intuitive since there must exist a set of conserved local charges τ_i^z preventing the transport of energy and other conserved quantities in the system, which underlie localization [26]. On the other hand, this construction is expected to hold only in the case that all eigenstates are localized, and the generalization is not immediate if the spectrum contains a many-body mobility edge.

The phenomenological model Eq. 1.20 predicts several features of the MBL phase that have been observed in numerical simulations and experiments. In fact, one of the earlier MBL studies considered a quench experiment where the entanglement entropy of a symmetric real space bipartition of the system in Eq. 1.18 was observed to grow logarithmically with time [27]. The initial states in the quench were product states of the physical spins. Additionally, the entanglement entropy saturated to an extensive value, though significantly smaller than the thermal value expected from the ETH.

The fact that the entanglement grows logarithmically follows from Eq. 1.20 [21]. Because of the form of the l-bit Hamiltonian, two l-bits i and k only exchange information through the direct interaction term that involves both l-bits, which is exponentially small in their distance. To see this, consider just three l-bits i, j, k, Ignoring the direct interaction between i and k, the Hamiltonian is $H = J_i^{(1)}\tau_i^z + J_j^{(1)}\tau_j^z + J_k^{(1)}\tau_k^z + J_{ij}^{(2)}\tau_i^z\tau_j^z + J_{jk}^{(2)}\tau_j^z\tau_k^z$. To understand the propagation of quantum information, it is useful to consider how an operator associated with an l-bit, e.g. $\tau_i^x = U\sigma_i^x U^\dagger$, changes in time according to the

Heisenberg picture for unitary quantum dynamics. In this picture, τ_i^x evolves in time as $\frac{d\tau_i^x(t)}{dt} = i[H, \tau_i^x(t)]$, where $\tau_i^x(0) = \tau_i^x$. The l-bit operator τ_i^x will generically spread out through the system. This spread in real space indicates that quantum information about the other l-bits has reached i and, thus, in general i becomes entangled with them. Because of the interaction with l-bit j, $\tau_i^x(t)$ evolves to terms that depend on l-bit j, namely $\tau_i^y \tau_j^z$, $\tau_i^x \tau_j^z$. However, since the interaction between j and k only depends on τ_j^z and commutes with the previous terms, information about k never reaches i through l-bit j. Thus, we see that l-bit i and k can only become entangled through direct interaction terms.

Therefore, two l-bits become entanglement after a time which is inversely proportional to their effective direct interaction, which scales exponentially with their distance R, i.e. $t \sim e^{R/\xi}$. Likewise, at time t, there is entanglement between every pair of l-bits within distance $R \sim \xi \ln(t)$ of each other, which underlies the observed logarithmic growth of entanglement entropy in quenches. This is in contrast with generic clean interacting systems including integrable systems and those that satisfy the ETH, where entanglement entropy grows linearly in time [28].

The l-bit picture also predicts certain properties of the eigenstates in the MBL phase. Because the quasi-local unitary U maps the product states of the physical spins $|s_1 \dots s_N\rangle$ to the eigenstates of the Hamiltonian, it is intuitive that entanglement statisfies an area law, as entanglement across the boundary is mostly due to the physical spins within distance ξ of the boundary. Thus, for $L \gg \xi$, the entanglement of a subregion does not increase with system size.

A consequence of the entanglement area law for eigenstates is that MBL systems cannot possibly satisfy the ETH. The latter predicts that the expectation value of local observables in eigenstates are identical to those of a thermal ensemble at the same energy, and thus the entanglement entropy of subsystems is generically extensive by definition. Similarly, the long time dynamics cannot be described by a thermal ensemble, since the local conserved charges τ_i^z also constrain the long time dynamics of the system.

1.3 Diagnostics

After the general overview of thermalization in quantum many-body systems, we now discuss some of the diagnostics, in addition to testing ETH directly, that are commonly used in numerical simulations to probe different aspects of thermalization and that are also useful for detecting its breakdown.

1.3.1 Level spacing statistics

Quantum systems are described by a Hamiltonian which defines a Hermitian matrix. Therefore, through diagonalization $H|E_m\rangle = E_m|E_m\rangle$ we obtain its spectrum formed by the sorted eigenenergies E_m, i.e. $E_{m-1} < E_m < E_{m+1}$. The spectrum contains some information about the dynamical properties of the system. For ergodic systems, the energy levels exhibit a phenomena known as level repulsion. Level repulsion means energy gaps between nearest neighbour energy levels, e.g. $s_m = E_m - E_{m-1}$, do not follow the expected statistical distribution if each level occurred randomly at a rate set by the average level spacing $\delta = \bar{s}$.

Under the assumption that levels occur randomly, given an eigenenergy E_m, the next energy level occurs with probability $\sim \frac{\Delta s}{\bar{s}}$ within any small energy window of range Δs. We would like to find the probability that the energy gap $s_n = E_{m+1} - E_m$ is in the range $[s, s + \Delta s]$. For this, we divide s in n small intervals of energy width Δs and demand that the next level does not turn up in any of these n intervals and that it appears in the next one. Therefore, this probability is given by $P(s) = (1 - \frac{\Delta s}{\bar{s}})^n \frac{\Delta s}{\bar{s}}$ with $n = \frac{s}{\Delta s}$. In the limit $\Delta s \to 0$, or equivalently $n \to \infty$, we find the so-called Poisson distribution $P_P(s) = \bar{s} e^{-\frac{s}{\bar{s}}}$. This probability density monotonically decreases with the size of the gap, i.e. it does not show level repulsion.

Because the Poisson distribution depends on the level spacing $\bar{s}$ which changes over the range of the spectrum, it is common practice in numerical simulations to measure level repulsion using the ratio of two neareast neighbour gaps $r_m = \frac{\min(s_m, s_{m-1})}{\max(s_m, s_{m-1})}$ [32], thus removing the dependency on the level spacing $\bar{s}$. Since in the absence of level repulsion the gaps s_m, s_{m-1} are uncorrelated, one can find from the gap distribution $P_P(s)$ the corresponding distribution for r given by $P_P(r) = \frac{2}{1+r^2}$.

It turns out the spectral rigidity of ergodic quantum many-body systems is well described by RMT [33, 45]. Depending on the symmetries of the Hamiltonian, the appropriate RMT ensemble is either the Gaussian Orthogonal Ensemble (GOE), the Gaussian Unitary Ensemble (GUE) or the Gaussian Symplectic Ensemble (GSE). For example, the specific ergodic models considered throughout this thesis contain anti-unitary symmetries related to time-reversal that constraint the Hamiltonian to be a real matrix in a specific basis, so their spectrum is well described by the GOE. In this ensemble, real symmetric matrices are sampled according to the probability distribution $P(H) \sim e^{-\mathrm{Tr}H^2}$, which is invariant with respect the action of orthogonal transformations O on H, i.e. $H \to O^T H O$. A very accurate ansatz for the distribution of ratios r_m for the RMT ensembles can be found through an exact calculation for 3×3 matrices [35]. Fig. 1.1 shows the distributions of ratios r for the Poisson distribution and for the GOE. The former decays with increasing

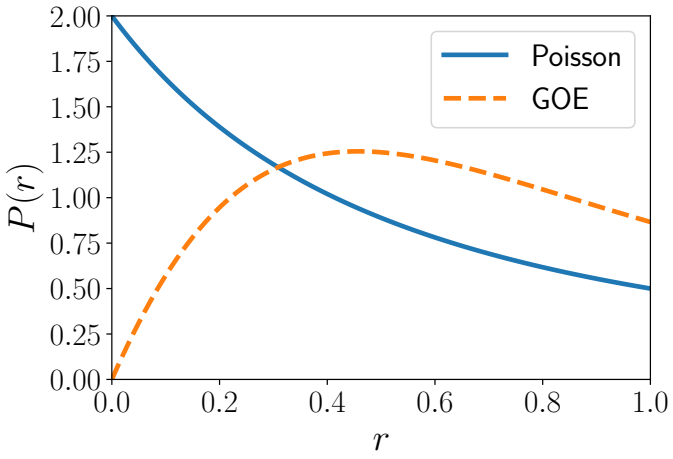

Figure 1.1: The distributions of the ratio of spectral gaps $r_m = \frac{\min(s_m, s_{m-1})}{\max(s_m, s_{m-1})}$ for Hamiltonians described either by Poisson statistics or sampled from the Gaussian Orthogonal Ensemble (GOE). The energy gaps $s_m = E_m - E_{m-1}$ are between the sorted eigenvalues $E_{m-1} < E_m < E_{m+1}$ of the Hamiltonian.

r, while the latter exhibits signatures of level repulsion as it peaks at a non-zero value of r. The two distributions can also be discriminated by their mean values $r_{\text{Poisson}} \approx 0.39$, $r_{\text{GOE}} \approx 0.53$ [35], which is a common diagnostic for level repulsion in numerical simulations.

A heuristic explanation for level repulsion in ergodic systems follows from considering the effect of a local perturbation on two nearest neighbour eigenstates. From the ETH ansatz Eq. 1.10, the off-diagonal matrix element of the local perturbation between the two levels $\sim 1/\sqrt{\rho}$ is asymptotically much larger than the level spacing $\sim 1/\rho$, since the density of states ρ increases exponentially with the number of degrees of freedom. From the application of degenerate perturbation theory, we see the two levels become separated by energy $\sim 1/\sqrt{\rho}$ which is much larger than the level spacing, so the levels repel each other.

1.3.2 Entanglement entropy of subsystems

The entropy of a pure state $|\psi\rangle$ describing a many-body system, defined as $S = -\text{Tr}\rho \log_2 \rho$ is identically zero, where $\rho = |\psi\rangle\langle\psi|$ is the density matrix of the full system [36]. Additionally, the entropy remains zero under unitary time evolution. Therefore, the connection

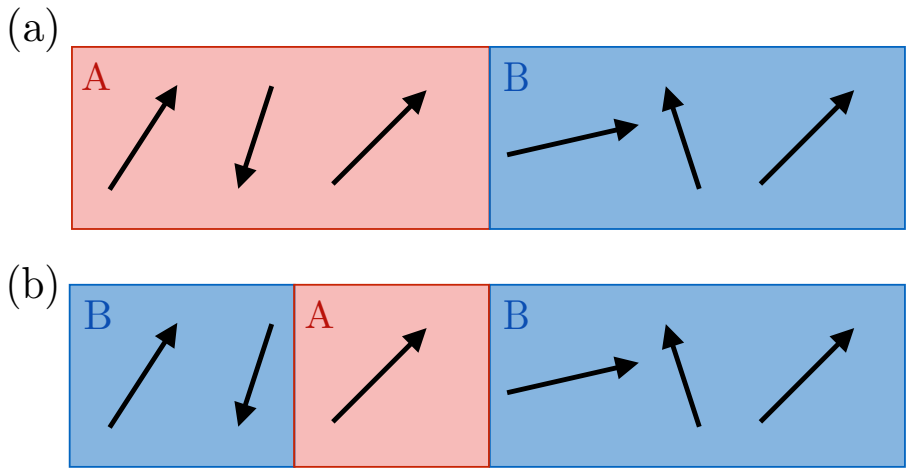

Figure 1.2: Two examples for bipartitions of an one dimensional chain formed by 1/2-spins: (a) symmetric half-chain bipartition, (b) bipartition defining the smallest possible subsystem composed by a single spin (subsystem A).

between entropy and thermalization requires the analysis of local subsystems. These become entangled under time evolution with other parts of the system, which effectively function as a heat bath.

Here we focus on subsystems which result from breaking the system into two parts A, B such that the total Hilbert Space $\mathcal{H}_{AB}$ is the direct product of the Hilbert Spaces of A and B, i.e. $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$. To illustrate this, we provide two examples for bipartitions in models describing one dimensional chains of 1/2-spin degrees of freedom. The first example corresponds to a symmetric bipartition of the chain, wherein subsystem A includes spins number 1 to $\frac{N}{2}$ (we assume an even number of 1/2-spins), while subsystem B includes the outstanding spins $N/2+1$ to N (see Fig. (a)). In the second example, subsystem A corresponds to a single spin, and subsystem B is composed by the other $N-1$ spins (see Fig. (b)).

A general pure state $|\psi\rangle$ reads $|\psi\rangle = \sum_{ij} \psi_{ij} |i\rangle_A |j\rangle_B$, where $|i\rangle_A$, $|i\rangle_B$ specify corresponding basis in subsystems A and B. As we shall see, the entanglement entropy of the subsystems A, B measures to what extent $|\psi\rangle$ is not separable, i.e. $|\psi\rangle \neq |\psi\rangle_A \otimes |\psi\rangle_B$.

A very useful tool for the purpose of analyzing the separability of subsystems A and B is the Schmidt decomposition [36], which we now describe. The coefficients ψ_{ij} can be interpreted as a matrix Ψ, where the number of rows n_A and columns n_B correspond

to the dimensions of the Hilbert spaces of the subsystems A and B, respectively. The singular value decomposition of Ψ, $\Psi = UsV^\dagger$, defines unitary matrices U of dimensions $n_A \times \min(n_A, n_B)$ and V of dimensions $n_B \times \min(n_A, n_B)$ and a diagonal matrix s of singular values of dimension $\min(n_A, n_B) \times \min(n_A, n_B)$.

As a result, the state $|\psi\rangle$ can be re-written in the following form:

$$|\psi\rangle = \sum_\gamma s_\gamma |\gamma\rangle_A \otimes |\gamma\rangle_B, \tag{1.21}$$

where $|\gamma\rangle_A = \sum_i U_{i\gamma}|i\rangle_A$ and $|\gamma\rangle_B = \sum_j V^\dagger_{\gamma j}|j\rangle_B$. Both sets of states $|\gamma\rangle_A$ and $|\gamma\rangle_B$ are orthonormal. Moreover, the singular values satisfy the following normalization condition: $\sum_\gamma s_\gamma^2 = 1$. It is thus natural to interpret $p_\gamma = s_\gamma^2$ as defining a probability distribution.

The reduced density matrix for subsystem A is defined as $\rho_A = \mathrm{Tr}_B \rho$, where $\rho = |\psi\rangle\langle\psi|$ is the density matrix for the full system. Similarly for subsystem B, $\rho_B = \mathrm{Tr}_A \rho$. Using Eq. 1.21, it follows immediately that $\rho_A = \sum_\gamma s_\gamma^2 |\gamma\rangle_{AA}\langle\gamma|$ and $\rho_B = \sum_\gamma s_\gamma^2 |\gamma\rangle_{BB}\langle\gamma|$.

Therefore, the state of the subsystems A, B is a statistical mixture, even when the global state is pure. Moreover, for any set of singular values s_γ distinct from the one where all the singular values are zero except for one, the state $|\psi\rangle$ is not separable and there is entanglement between the two subsystems.

The entanglement entropy of ρ_A is defined as $S_A = -\mathrm{Tr}\rho_\mathrm{A} \log \rho_\mathrm{A}$, and it is identical to the classical Shannon entropy of the probability distribution $p_\gamma = s_\gamma^2$. Consequently, S_A measures the logarithm of the effective number of states in A contributing to the statistical mixture in ρ_A. When the full system is in a pure state, it follows from the Schmidt decomposition Eq. 1.21, that both subsystems have the same entanglement entropy, i.e. $S_A = S_B$. The entanglement entropy of a subsystem takes values between zero and the logarithm of the dimension of the Hilbert Space of the subsystem. The minimum value $S_A = 0$ occurs when $|\psi\rangle$ is separable and the maximum value $S_A = \log(n_A)$ occurs when when ρ_A is in a maximally mixed state defined by $\rho_A = 1/n_A$ which is proportional to the identity matrix.

The ETH ansatz for the expected values and fluctuations of local observables evaluated in eigenstates of ergodic quantum many-body systems provides non-trivial constraints on the entanglement entropy of subsystems. More specifically, the ETH ansatz predicts that for eigenstates the reduced density matrix of local subsystems converges with increasing system size to the thermal density matrix $\rho_A^\beta \sim e^{-\beta H_A}$ of the subsystem. H_A is the restriction of the Hamiltonian H to subsystem A and the inverse temperature $\beta = 1/T$ is set by the energy of the eigenstate. Since $\rho_A \approx \rho_A^\beta$, it immediately follows (by definition) that $S_A = S_\mathrm{Thermal}$, where S_Thermal is the thermodynamic entropy of subsystem A at the

corresponding energy density. Accordingly, S_A is proportional to the volume of subsystem A and thus satisfies a volume law, which follows from the extensivity of the thermodynamic entropy. Conversely, if for a quantum many-body system the scaling of S_A with subsystem size does not follow a volume law, it implies the breakdown of the ETH and thermalization, which is the case for MBL systems.

We end this subsection with a comment on the calculation of the entanglement entropy of a subsystem A in numerical simulations. The global state $|\psi\rangle$ defines a vector that can be reshaped into a matrix Ψ (as defined above). Further, the Singular Value Decomposition of Ψ provides the set of singular values s_γ. Alternatively, from the matrix Ψ, the reduced density matrix of subsystem A can be obtained directly by a matrix multiplication $\rho_A = \Psi\Psi^\dagger$; the eigenvalues of ρ_A are identical to s_γ^2 and can be obtained by exact diagonalization. Finally, the entanglement entropy of subsystem A is given by $S_A = -\sum_\gamma s_\gamma^2 \log(s_\gamma^2)$.

1.3.3 Participation Ratios

The scaling of eigenstate entanglement entropy is now a standard diagnostic for thermalization in quantum many-body systems. Similarly, eigenstates can be characterized by their participation ratio in some basis, which effectively measures how many states in this basis contribute to the eigenstate. Since the ETH is motivated by RMT, where eigenvectors are essentially random in any local basis, eigenstates of thermalizing quantum many-body Hamiltonians are expected to be delocalized in the Hilbert Space. For the 1/2-spin models considered in this thesis, an example of a local basis are the product states $|s_1 s_2 \dots s_N\rangle$, where for each spin $s_i = \langle\sigma_i^z\rangle = \pm 1$.

Concretely, for any eigenstate $|E_m\rangle$, its participation ratio on a basis $|\gamma\rangle$ is defined as:

$$P_m = \frac{1}{\sum_\gamma |\langle\gamma|E_m\rangle|^4} \tag{1.22}$$

The participation ratio P_m ranges from 1, in the case of only one non-zero amplitude $\langle\gamma|E_m\rangle$, to the dimension of the Hilbert Space $\mathcal{D}$, when all the amplitudes have the same norm equal to $1/\sqrt{\mathcal{D}}$.

For random states, the expected value of the participation ratio is equal to 1/3 of the dimension of the Hilbert Space $\mathcal{D}$ [37]. This value is in close agreement to what is observed for eigenstates of ergodic systems at infinite temperature, showing that these states are indeed essentially random on a local basis. At other energy densities, the participation ratios remain a finite fraction of the total Hilbert Space dimension [37].

Therefore, if a quantum many-body system satisfies ETH, the participation ratios of eigenstates on a local basis grow exponentially with the number of degrees of freedom.

1.4 Periodic driving

Quantum systems coupled to periodic time-varying external fields are ubiquitous in the study of physics. Examples include materials irradiated by periodic fields, their response to periodic currents or mechanical shaking, etc [46]. These apparently unexceptional phenomena resulted in notable applications across multiple domains.

The impact of the laser in Physics has been significant: it has allowed the creation of synthetic matter using optical lattices [47], and more than 100 years after the theoretical prediction of Einstein, the Laser Interferometer Gravitational-Wave Observatory (LIGO) finally made the experimental verification of gravitational waves [48]. The laser also led to breakthroughs in other fields such as technology (e.g. optical drives) and medicine (e.g. eye surgery).

Another remarkable example is Nuclear Magnetic Resonance [49] in which atomic nuclei interact with time periodic magnetic fields. The NMR protocol has not only been applied to collect structural information on molecules, crystals and other physical structures, but is also used to perform brain imaging through the detection of local variations in blood flow, which has many applications in medicine and neuroscience.

In condensed matter, the recent interest in these systems was sparked by the discovery that periodic driving can induce topological properties in the band-structures of materials. For example, it was shown that radiating circularly polarized light in graphene generates topological edge states [50].

1.4.1 The kicked rotor

At the single-particle level, periodic-driving already reveals unexpected phenomena. A well-known example is the quantum kicked rotor [38, 39, 40, 41, 42, 43, 44, 45]. This system corresponds to a particle which moves freely, but every period is *kicked* by a potential. This is described by the Hamiltonian $H = p^2/2m + k\cos(x)\sum_{n=1}^{\infty}\delta(t-nT)$.

Classically, the kick changes the momentum of the particle by a finite amount $\sim k\sin(x)$ every period, which depends on the position x of the particle and can have arbitrary sign. As a result, we observe random walk behavior for strong enough kicks; the momentum of the particle undergoes diffusive motion growing as $p^2 \sim n$, where n is the number of elapsed periods.

Surprisingly, the quantum kicked rotor behaves quite differently than the classical model and the momentum of the particle remains finite in the long time limit. This phenomenon

was dubbed dynamical localization, since it is reminiscent of Anderson localization. However, in this case, it happens in momentum space. Interestingly, the two phenomena are related. The kicked rotor can be mapped to a time independent problem analogous to Anderson localization except for a few modified ingredients that do not change the physical properties of the system (e.g. the on-site energies in this case are only pseudorandom).

1.4.2 The Floquet Operator

More generically, a periodically-driven quantum system is defined by a periodic time dependent Hamiltonian $H(t)$ which satisfies $H(t+T) = H(t)$, where T is the period of the drive. The solutions to the Schrodinger equation under a time dependent Hamiltonian satisfy:

$$H(t)|\psi(t)\rangle = i\frac{d|\psi(t)\rangle}{dt}. \tag{1.23}$$

If one is only interested in the stroboscopic dynamics, i.e. in times $t = nT$ with n integer (the beginning of the period can be chosen arbitrarily), or in the infinite time dynamics, then this information is encoded in the evolution operator over a single period of the drive, the Floquet operator. For a generically time dependent periodic Hamiltonian, the Floquet operator reads:

$$F = \mathcal{T}\int_0^T \exp(-iH(t)t)\,dt. \tag{1.24}$$

Here $\mathcal{T}$ denotes the time ordering of the products of the Hamiltonian $H(t_1)\dots H(t_n)$ arising from the expansion of the exponential.

Since F is a unitary operator, it can be expressed in terms of a Hermitian operator H_F that naturally plays the role of an effective Hamiltonian and is dubbed the Floquet Hamiltonian:

$$F = e^{-iH_F T}. \tag{1.25}$$

This follows from the fact that a unitary operator can always be diagonalized and additionally its eigenvalues are complex numbers of norm 1 due to unitarity:

$$1 = \langle\psi_i|\psi_i\rangle = \langle\psi_i|F^\dagger F|\psi_i\rangle = \lambda_i^*\lambda_i, \tag{1.26}$$

where $\lambda_i, |\psi_i\rangle$ denote an eigenvalue of F and the associated eigenvector, respectively. Consequently, in this eigenstate basis, the Floquet operator has the following form:

$$F = \sum_i e^{-i\omega_i T}|\psi_i\rangle\langle\psi_i|. \tag{1.27}$$

Thus, the Floquet Hamiltonian is defined as

$$H_F T = i \ln F = \sum_i \omega_i T |\psi_i\rangle\langle\psi_i|. \tag{1.28}$$

Since the quasi-energies ω_i are only defined modulo $2\pi/T$, the Floquet Hamiltonian is not unique.

Eq. 1.25 implies that with respect to the stroboscopic dynamics the system is effectively described by a time independent Hamiltonian H_F. However, there are no guarantees that this Hamiltonian is physical. It might be non-local, for instance.

1.4.3 Periodic Gibbs Ensemble

The existence of a Floquet Hamiltonian has interesting consequences for non-interacting quantum many-body systems described by periodic Hamiltonians that at each time remain quadratic in terms of single-particle creation and annihilation operators. The equilibrium state reached by these systems after a quench experiment is now well understood [51].

In this setting, we can show that the Floquet Hamiltonian H_F is also quadratic in the single-particle creation and annihilation operators and, consequently, can be cast in a diagonal form $H_F = \sum_m E_m a_m^\dagger a_m$. Therefore, the particle numbers $a_m^\dagger a_m$ commute with the Floquet Hamiltonian H_F and as a result are conserved by the stroboscopic dynamics of the system.

In analogy with the Generalized Gibbs Ensemble (see Section 1.1.2), we can define the ensemble with fixed expectation values of the conserved quantities $a_m^\dagger a_m$ that maximizes the entropy. It is referred to as the Periodic Gibbs Ensemble [51] and has the following form:

$$\rho_{\mathrm{PGE}} = e^{-\sum_m \ell_m a_m^\dagger a_m}, \tag{1.29}$$

where the Lagrange multipliers ℓ_m are determined by the expectation value of the conserved quantities $a_m^\dagger a_m$ in the initial state of the system.

Indeed, numerical simulations [51] have shown that this ensemble provides an accurate description of the long time properties of these systems after equilibration takes place.

1.4.4 Magnus expansion

The definition of the Floquet Hamiltonian $H_F = \frac{-i}{T} \ln \left(\mathcal{T} \int_0^T \exp(-iH(t)t)\, dt \right)$ for generic interacting many-body Hamiltonians does not allow for a general and tractable solution.

However, it is possible to determine it perturbatively in the driving period T (on the scale of the local energy of $H(t)$ which we define by J) using the so-called Magnus expansion [53, 54, 52], which is derived in the following.

We recast the Schrodinger equation in the form: $\frac{d|\psi(t)\rangle}{dt} = A(t)|\psi(t)\rangle$, with $A(t) = -iH(t)$. We would like to find $\Omega(t)$ such that $|\psi_0(t)\rangle = e^{\Omega(t)}|\varphi_0\rangle$, where $|\varphi_0\rangle$ is the state at time $t = 0$. In this notation, the Schrodinger equation is equivalent to

$$\frac{de^{\Omega(t)}}{dt}e^{-\Omega(t)} = A(t). \tag{1.30}$$

Using the Poincare-Hausdorff matrix identity $\frac{de^{\Omega}}{dt} = e^{\Omega}\frac{1-e^{-\mathrm{ad}_\Omega}}{\mathrm{ad}_\Omega}$, where ad_Ω corresponds to the mapping $B \to [\Omega, B]$ for an general operator B, and the identity $e^{-\Omega}Ae^{\Omega} = e^{-\mathrm{ad}_\Omega}A$, we find:

$$\frac{d\Omega}{dt} = \frac{\mathrm{ad}_\Omega}{e^{\mathrm{ad}_\Omega} - 1}A. \tag{1.31}$$

From Eq. 1.31, a recursive expansion of $\Omega(t)$ in powers of A can be obtained: $\Omega(t) = \sum_{i=1}^{\infty}\Omega^{(n)}(t)$, where $\Omega^{(n)}$ is of order $||A||^n$. Note that by means of the identity $\Omega(T) = -iH_FT$, we obtain an expansion for the Floquet Hamiltonian $H_F = \sum_{n=1}^{\infty}H_F^{(n)}$ as well.

To first order in A, Eq. 1.31 reads $\frac{d\Omega^{(1)}(t)}{dt} = A(t)$, which implies $\Omega^{(1)}(t) = \int_0^t dt_1 A(t_1)$ or equivalently $H_F^{(1)} = \frac{1}{T}\int_0^T dt' H(t')$ given that $A(t) = -iH(t)$. In the large frequency limit, it is indeed intuitive that the system is described by the time averaged Hamiltonian, since the system's own dynamical timescale $\sim 1/J$ is much slower than the driving.

To second order in A, we find $\frac{d\Omega^{(2)}(t)}{dt} = -\frac{1}{2}[\Omega^{(1)}(t), A(t)]$, which is solved in terms of $A(t)$ given that $\Omega^{(1)}(t) = \int_0^t dt_1 A(t_1)$. As a result, the second order correction to the Floquet Hamiltonian is $H_F^{(2)} = \frac{1}{2T}\int_0^T dt_1 \int_0^{t_1} dt_2 [H(t_1), H(t_2)]$. This procedure can be iterated to obtain $\Omega(t)$ up to the desired order in $A(t)$.

It is instructive to also present the third order contribution to the Floquet Hamiltonian:

$$H_F^{(3)} = -\frac{1}{6T}\int_0^T dt_1 \int_0^{t_1} dt_2 \int_0^{t_2} dt_3 \left([H(t_1), [H(t_2), H(t_3)]] + [H(t_3), [H(t_2), H(t_1)]]\right). \tag{1.32}$$

The pattern for $n = 1, 2, 3$ suggests that $H_F^{(n)}$ depends on terms with n nested commutators involving the Hamiltonian $H(t)$. Because $H(t)$ is assumed to be local, at each order $H_F^{(n)}$

is of order $\frac{1}{T}(TJ)^n$. Consequently, we see the Mangus expansion corresponds to a small period expansion of the Floquet Hamiltonian.

Moreover, at each order, $H_F^{(n)}$ is still a local Hamiltonian, but the range of the local terms increases linearly with n. To illustrate this point, we consider the example where $H(t)$ describes a 1/2-spin chain with only time dependent nearest neighbour couplings, e.g. $\sigma_i^x\sigma_{i+1}^x$, $\sigma_i^y\sigma_{i+1}^y$. At second order, in the definition of $H_F^{(2)}$, most of the terms in $[H(t_1), H(t_2)]$ commute with each other, except overlapping operators, such as $[\sigma_1^x\sigma_2^x, \sigma_2^y\sigma_3^y] = 2i\sigma_1^x\sigma_2^z\sigma_3^y$. Hence, the range of the Floquet Hamiltonian H_F at second order has grown compared to the time averaged Hamiltonian assumed to only have nearest neighbour couplings. By an analogous argument, $H_F^{(n)}$ contains local terms with range $\sim n$.

We now explore the consequences of the convergence of the Magnus expansion. Here the term convergence is used in the sense that truncating the series at order $O(L^0)$ provides an accurate description of the stroboscopic dynamics (i.e. at times nT) under H_F. In this situation, H_F is essentially a sum of local terms. This point has important consequences, in particular if $H(t)$ is ergodic. If H_F is local, we are back to the usual setting where the system is described by a time independent local ergodic Hamiltonian. Consequently, the long time steady state after a quench protocol from an initial state $|\varphi_0\rangle$ is described by the thermal state $\rho = \frac{e^{-\beta H_F}}{Z}$, where the inverse temperature $\beta = 1/T$ is fixed by the energy of the initial state $\langle\varphi_0|H_F|\varphi_0\rangle$.

On the other hand, if H_F is of infinite range, then no local conserved quantities constrain the dynamics of the system. Consequently, any initial state equilibrates to what is effectively an infinite temperature state $\rho_\infty = \frac{1}{Z}$, which maximizes the entropy under no constraints. When the system equilibrates to the infinite temperature ensemble ρ_∞, this is usually referred to as *heating up to infinite temperature*.

The only mathematical guarantee for the convergence of the series is if $\int_0^T ||H(t)||dt \lesssim O(1)$ [54], where the norm refers to the matrix norm. This result provides no insight for generic local interacting quantum many-body systems, since their spectrum is extensive in the number of degrees of freedom, which means the previous bound is eventually violated for a large enough system.

1.4.5 Replica formalism

We end this overview of periodic driving with a discussion of one of the earliest strategies to study periodically driven systems, the so-called replica formalism [55]. This approach transforms the time dependent problem into an effectively time independent one, but at

the cost of enlarging the dimension of the Hilbert space. Recall that the solutions to the Schrodinger equation under a time dependent Hamiltonian satisfy:

$$H(t)|\psi(t)\rangle = i\frac{d|\psi(t)\rangle}{dt}. \tag{1.33}$$

According to the Floquet theory, there is a complete set of solutions of the form $|\psi(t)\rangle = e^{-i\varepsilon t}|\varphi(t)\rangle$, where the state $|\varphi(t)\rangle$ has period T and ε is called a quasi-energy. Notice the presence of degenerate solutions: if $|\varphi(t)\rangle$ is a solution, so is $|\varphi(t)\rangle e^{in\omega t}$, but the associated quasi-energy also shifts: $\varepsilon \to \varepsilon - \omega n$. For a solution of this form, Eq. 1.33 becomes:

$$\left[H(t) - i\frac{d}{dt}\right]|\varphi(t)\rangle = \varepsilon|\varphi(t)\rangle \tag{1.34}$$

We now show that Eq. 1.34 is equivalent to a time independent Schrodinger equation. To that purpose, notice that $|\varphi(t)\rangle = \sum_\alpha A_m(t)|E_m\rangle$, where the states $|E_m\rangle$ constitute some basis of the Hilbert space, which we can choose to be the basis defined by the eigenstates of the time averaged Hamiltonian $H(0) = \frac{1}{T}\int_0^T H(t)dt$. Since the amplitudes $A_m(t)$ have period T, they can be expanded in a Fourier series $A_m(t) = \sum_{k=-\infty}^{\infty} A_m^{(k)} \frac{1}{\sqrt{T}} e^{ik\omega t}$, and as consequence:

$$|\varphi(t)\rangle = \sum_{n,\alpha} A_m^{(k)} \frac{1}{\sqrt{T}} e^{ik\omega t}|E_m\rangle. \tag{1.35}$$

Eq. 1.35 clearly indicates that $|\varphi(t)\rangle$ can be interpreted as belonging to a larger Hilbert space spanned by the orthonormal basis $|E_m, k\rangle \equiv |E_m\rangle \otimes |k\rangle$. Here, $|k\rangle = \frac{1}{\sqrt{T}} e^{ik\omega t}$ with $n \in \mathbb{Z}$ and the inner product for these monochromatic states is defined as $\langle k|l\rangle = \frac{1}{\sqrt{T}}\int_0^T e^{-i(k-l)\omega t} = \delta_{k,l}$. Thus, in this formulation, Eq. 1.34 is equivalent to the time independent problem:

$$\sum_{l,n} \langle E_m|H(k-l)|E_n\rangle A_n^{(l)} = (\varepsilon - k\omega)A_m^{(k)}, \tag{1.36}$$

where the $H(k) = \frac{1}{T}\int_0^T dt H(t)e^{-ik\omega t}$ denote the Fourier components of the Hamiltonian.

In the absence of driving, i.e. $H(k) = 0$ for $k \neq 0$, the quasi-energies are of the form $\varepsilon_{m,k} = E_m + k\omega$, where E_α are the energies of the eigenstates of the static Hamiltonian $H(0)$. If a driving protocol is turned on, i.e. terms $H(k)$ with $k \neq 0$ become non-zero, the eigenstates of Eq. 1.36 will generically consist of a mixture of different eigenstates of $H(0)$. However, there's a stricking difference in this setting compared to the situation where a time

independent perturbation is added to $H(0)$. While in the latter only eigenstates of $H(0)$ with similar energy are nearly degenerate, in the periodic-driving setting, eigenstates whose energy differs approximately by a multiple of ω are also nearly degenerate. Specifically, if $E_n - E_m \sim l\omega$, then the states $|E_m, k+l\rangle$ and $|E_n, k\rangle$ are nearly degenerate $\varepsilon_{m,k+l} \sim \varepsilon_{n,k}$ and can be mixed by the Fourier mode $H(l)$ of the Hamiltonian. Hence, the time periodic setting introduces new ways to delocalize the eigenstates of the time independent Hamiltonian $H(0)$ and it is natural to expect new dynamical regimes to emerge as a consequence of that.

Chapter 2

Periodically-driven ergodic and many-body localized quantum systems: local drive

2.1 Introduction

The recent interest in periodically driven quantum many-body systems with local interactions started with D'Alessio and Polkovnikov [56] who hypothesized that two distinct dynamical regimes were possible: the system either (i) keeps absorbing energy, heating up to infinite temperature (e.g. defined using the time averaged Hamiltonian) at long times or (ii) dynamically localizes at a certain energy similar to the case of the kicked rotor.

Here we establish the conditions under which the dynamics of locally driven many-body systems are described by a non-local versus a local Floquet Hamiltonian H_F (defined in Section 1.4.2), which result respectively in the dynamical regimes (i) and (ii). Specifically, we consider the following driving protocol illustrated in Fig. 2.1(a), where the Hamiltonian is switched between two operators periodically in time,

$$H(t) = \begin{cases} V & 0 < t < T_1 \\ H_0 & T_1 < t < T_0 + T_1. \end{cases} \tag{2.1}$$

$H_0(V)$ is applied during time T_0 (T_1), such that the total period $T = T_0 + T_1$. We will consider lattice systems, in which the Hilbert space on every site is finite dimensional,

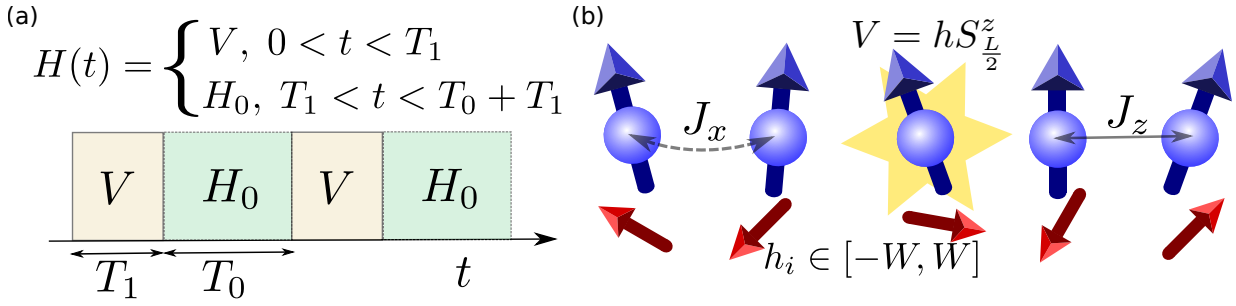

Figure 2.1: (a) A scheme of the general driving protocol. (b) Example of a 1d XXZ spin chain studied numerically (blue arrows). The Hamiltonian H_0 contains nearest-neighbor hopping and interactions, in the presence of random z field (red arrows). Driving is performed locally by the local operator $V = hS^z_{L/2}$ applied to the middle spin.

the operator H_0 corresponds to a Hamiltonian describing a quantum-many body system with short range interactions, while V is a local operator that only acts on a finite set of contiguous degrees of freedom.

This chapter is organized as follows. In Section 2.2, we map the spectral problem for the Floquet operator describing the system onto an effective hopping problem and show that the competition between the typical off-diagonal matrix elements of $\tan(VT_1/2)$ with respect to the eigenstates of H_0 and the typical level spacings in the spectrum of H_0 determines whether or not the system will heat up to infinite temperature at long times. In Section 2.3, we discuss some of the signatures of heating in the hopping problem and introduce the spin model that is studied numerically. In Sections 2.4.1 and 2.4.2, we then apply our criterion to two general classes of driven many-body systems. The first class is ergodic systems, i.e. systems that act as their own heat bath and satisfy the Eigenstate Thermalization Hypothesis [9, 10, 7], c.f. Section 1.1.2. In Section 2.4.1, we show that for ergodic systems, the system heats up to infinite temperature under driving and thus the Magnus expansion in the driving period T diverges and H_F is unphysical. The second class is many-body localized systems with quenched disorder that are known to be non-ergodic [23, 30, 32, 25, 26], c.f. Section 1.2. Under local driving, we show in Section 2.4.2 that MBL systems retain memory of their initial state and never reach infinite temperature. Here the Magnus expansion is expected to converge, H_F is local and itself MBL. Throughout Section 2.4, we support our analytical results with numerical studies of the driven XXZ 1/2-spin chain in random z fields. Our conclusions are summarized in Section 2.5.

2.2 Mapping the Floquet onto a hopping problem

In this Section, we provide a mapping of the Floquet problem onto an effective hopping model, which allows us to infer the properties of H_F, without explicitly computing and analysing the properties of the Magnus expansion. This mapping shows that the competition between the typical level spacing and the hopping matrix element (i.e., the matrix element of $\tan(VT_1/2)$ between the eigenstates of H_0) determines the structure of the Floquet eigenstates.

The Floquet operator for the driving protocol in Fig. 2.1(a) is given by

$$F = \exp(-iH_0T_0)\exp(-iVT_1). \tag{2.2}$$

The eigenstates of F completely determine the stroboscopic evolution of the system. Below, we map the eigenvalue problem of F onto a hopping problem, similar to the kicked rotor model in Ref. [57]. The lattice sites of the hopping problem represent eigenstates of H_0, while V induces hopping between sites.

The Floquet operator is unitary and its spectrum is given by (c.f. Section 1.4.2):

$$F|\psi_i\rangle = e^{-iH_FT}|\psi_i\rangle = e^{-i\omega_iT}|\psi_i\rangle,\ i = 1,\ldots,\mathcal{D}, \tag{2.3}$$

where $\mathcal{D}$ is the dimensionality of the Hilbert space (e.g., $\mathcal{D} = 2^L$ for the system of L 1/2-spins considered below) and $\langle\psi_i|\psi_j\rangle = \delta_{ij}$. The quasi-energies ω_i are defined modulo $2\pi/T$, hence H_F is not unique.

Finding the spectrum of F in a many-body system is generally hard, since F is generically highly non-local. For example, in the case of a 1/2-spin chain, F can be expressed in the basis of Pauli matrices as $F = \sum_{i,\alpha} f_i^\alpha \sigma_i^\alpha + \sum_{ij,\alpha\beta} f_{ij}^{\alpha\beta}\sigma_i^\alpha\sigma_j^\alpha + \ldots$, and we expect the weights $f_{\{i\}}^{\{\alpha\}}$ to be non-zero for terms where the set $\{i\}$ includes pairs of sites whose distance is of the order of the system size L.

To circumvent this difficulty, let us provide an explicit mapping to a local Hamiltonian problem. Rewrite e^{-iVT_1} in terms of a Hermitian operator G as:

$$e^{-iVT_1} = (1+iG)(1-iG)^{-1},\ G = -\tan(VT_1/2). \tag{2.4}$$

In general, G is not spatially local in the same sense that F is not local, as just explained above. If, however, V is local, then G is also local. The diagonalization problem (Eq. 2.3) is then equivalent to:

$$e^{-iH_0T_0}\frac{1+iG}{1-iG}|\psi_i\rangle = e^{-i\omega_iT}|\psi\rangle \tag{2.5}$$

Defining $|\chi_i\rangle \equiv (1-iG)^{-1}|\psi_i\rangle$, Eq. 2.5 becomes:

$$e^{-iH_0T_0}(1+iG)|\chi_i\rangle = e^{-i\omega_i T}(1-iG)|\chi_i\rangle \tag{2.6}$$

$$-i\frac{e^{-iH_0T_0}-e^{-i\omega T}}{e^{-iH_0T_0}+e^{-i\omega_i T}}|\chi_i\rangle = G|\chi_i\rangle \tag{2.7}$$

And finally, using the identity $\tan(x) = -i\frac{e^{2ix}-1}{e^{2ix}+1}$, we obtain:

$$\left[\tan\frac{1}{2}(H_0T_0-\omega_i T) - G\right]|\chi_i\rangle = 0. \tag{2.8}$$

Let us view the eigenbasis of H_0, labeled by $|E_m\rangle$, as sites in a lattice. Solving Eq. 2.8 is equivalent to finding the zero-energy eigenstate of a hopping problem on this lattice, where $\lambda_m(\omega_i) = \tan\frac{E_mT_0-\omega_iT}{2}$ plays the role of an on-site energy on site m, and G_{nm} is the hopping amplitude between sites n and m. The competition between the typical level spacing and the hopping matrix element, which is different in the ergodic and MBL phases [11, 12, 14, 30], thus determines the structure of the Floquet eigenstates.

2.3 Microscopic model and heating diagnostics

We now define several quantities that can serve as a measure for when the system heats up under periodic driving. We also introduce the model of a disordered XXZ spin chain, where these quantities can be readily computed in numerical simulations. These results are presented in Section 2.4.

We imagine preparing the system at $t = 0$ in a low-energy eigenstate of H_0, $|\varphi_0\rangle$. The stroboscopic evolution at times $t = nT$ follows from the expansion of $|\varphi_0\rangle$ in the Floquet eigenbasis:

$$|\varphi_n\rangle = F^n|\varphi_0\rangle = \sum_i A_i \exp(-i\omega_i nT)|\psi_i\rangle, \tag{2.9}$$

where $A_i = \langle\psi_i|\varphi_0\rangle$. At long times, the time averaged density matrix is

$$\rho_\infty = \sum_i |A_i|^2|\psi_i\rangle\langle\psi_i|. \tag{2.10}$$

The nature of the eigenstates $|\psi_i\rangle$ determines the steady state as $t \to \infty$. If each $|\psi_i\rangle$ is delocalized in the eigenbasis of H_0, then each $|\psi_i\rangle$ corresponds to an infinite temperature state. The entire density matrix, ρ_∞, describes a system at infinite temperature in this

case. If on the other hand the $|\psi_i\rangle$ are localized in the eigenbasis of H_0, then, depending on A_i, ρ_∞ describes a system at different energies.

To characterize the energy absorbed under driving, at each $t = nT$ we introduce a dimensionless energy

$$Q_n = \frac{(\langle\varphi_n|H_0|\varphi_n\rangle - E_0)}{(E_{T=\infty} - E_0)}, \tag{2.11}$$

where $E_{T=\infty} = \mathrm{Tr}(H_0)/\mathcal{D}$ is the average energy at infinite temperature, and E_0 is the energy at $t = 0$. As we shall see, the cases where H_0 is ergodic or MBL are distinguished by $Q_n \to 1$ and $Q_n \to 0$ as $n \to \infty$, respectively.

Finally, to quantify the structure of the Floquet eigenstates, we compute the participation ratio (PR). For the Floquet eigenstate $|\psi_i\rangle$, PR in the basis of eigenstates $|E_m\rangle$ is defined as

$$\mathrm{PR} = \left(\sum_m |A_{mi}|^4\right)^{-1}, \tag{2.12}$$

where $A_{mi} = \langle E_m|\psi_i\rangle$.

The above quantities can be readily computed in finite spin chains using exact diagonalization. For concreteness, we focus on the XXZ 1/2-spin chain with L sites and open boundary conditions, illustrated in Fig. 2.1:

$$H_0 = J_x \sum_{i=1}^{L-1}(S_i^x S_{i+1}^x + S_i^y S_{i+1}^y) + J_z \sum_{i=1}^{L-1} S_i^z S_{i+1}^z + \sum_{i=1}^{L} h_i S_i^z, \tag{2.13}$$

where fields h_i are independent random variables drawn from the uniform distribution $[-W, W]$, and we fix $J_x = J_z = 1$. The model described by Eq. 2.13 exhibits both ergodic and MBL phases as a function of disorder strength W, with the transition at $W_* \approx 3$ [30]. The system is driven by

$$V = hS_{L/2}^z \tag{2.14}$$

acting on the middle spin (we assume $h = 2$). Since for finite system sizes the range $\Delta \sim O(L)$ of the spectrum of H_0 is also finite, in order to minimize finite size effects we choose $T_0\Delta \gg 1$ and we fix $T_0 = 7$. This guarantees that the initial state $|\varphi_0\rangle$ can find resonant states at energy difference $\pm\omega = \pm 2\pi/T_0$. The results for this model will be presented in Section 2.4 for both the ergodic and MBL cases, which correspond to disorder strength $W = 0.5$ and $W = 8$, respectively. The number of disorder samples analyzed ranges from 2×10^4 ($L = 8, 10$) to $\sim 10^3$ ($L = 12, 14$).

2.4 Local driving

Here we derive the conditions for the system to heat up to infinite temperature, and confirm them numerically. We will assume that V is local (e.g. as in Eq. 2.14). Then, $G = -\tan\frac{VT_1}{2}$ is also local. Furthermore, we will assume that the Hilbert space at every lattice site is finite dimensional, like in the model described by Eq. 2.13, such that G has a finite matrix norm, $||G|| \leq C$, with the possible exception of special values of T_1 that verify $\eta_i T_1 = (2n+1)\pi$ for an eigenvalue η_i of V and arbitrary integer n. Notice that for the model described by Eqs. 2.13 and 2.14 the term e^{-iVT_1} becomes the identity for these values of T_1 and the Floquet Hamiltonian is identical to H_0.

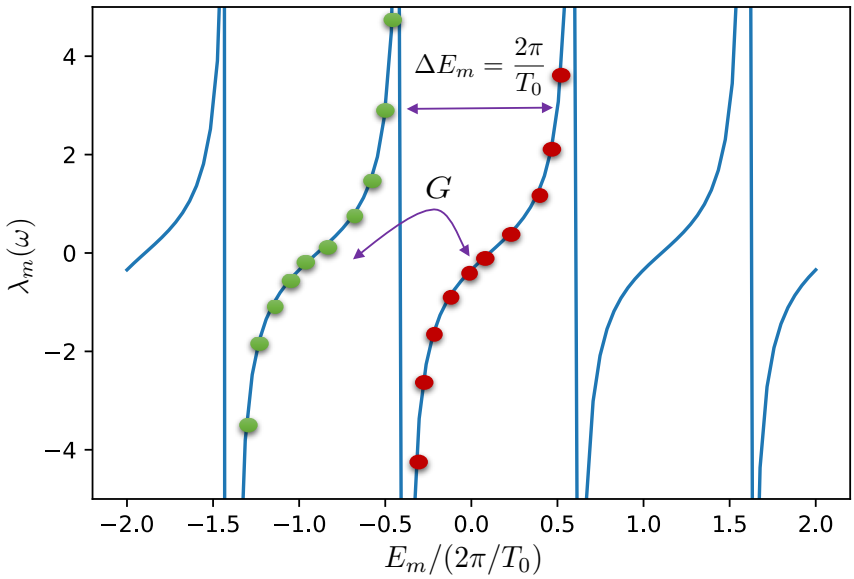

Figure 2.2: *On-site* energy $\lambda_m(\omega) = \tan\frac{E_m T_0 - \omega T}{2}$ for the eigenvalue problem associated with the spectrum of the Floquet Operator F (see Eq. 2.8); only the eigenstates belonging to the two central bands are shown for simplicity (represented by green and red dots). Eigenstates of H_0 which differ in energy E_m by a finite amount $2\pi/T_0$ are almost degenerate in terms of the energy $\lambda_m(\omega)$. The driving operator G couples these states with strength G_{mn} and if this is larger than the level spacing $\Delta\lambda$ these states strongly mix. In this case, we would expect the eigenstates $|\chi_i\rangle$ to be completely delocalized over the eigenstates of H_0.

We now explore the properties of the zero-energy states $|\chi_i\rangle$ in Eq. 2.8, to which the

eigenstates of the Floquet Operator $|\psi_i\rangle$ are related by the action of a local operator: $|\psi_i\rangle = (1 - iG)|\chi_i\rangle$.

We expect the states $|\chi_i\rangle$ to be localized (delocalized) on the eigenbasis of H_0 when the typical hopping amplitude G_{mn} between eigenstates of H_0, $|E_m\rangle$, $|E_n\rangle$ belonging to neighbouring bands which differ in energy $|E_m - E_n|$ by $\sim \frac{2\pi}{T_0}$ (see Fig. 2.2) is much smaller (larger) than the level spacing $\Delta\lambda$ between them; this follows from a first order perturbation theory analysis on G determining if the states $|E_m\rangle$ and $|E_n\rangle$ hybridize or not.

At quasi-energy ω (defining the eigenvalue of F), the on-site energies $\lambda_m(\omega) = \tan\frac{E_m T_0 - \omega T}{2}$ have a distribution $P(\lambda) \propto \frac{1}{1+\lambda^2}$ if the E_m are distributed uniformly. Because $|\chi_i\rangle$ satisfies Eq. 2.8, the expected value of the operator $\lambda(\omega) = \tan\frac{H_0 T_0 - \omega T}{2}$ for this state is $\bar{\lambda} = \langle\chi_i|G|\chi_i\rangle$. This suggests that the relevant values of the on-site energies $\lambda_m(\omega)$ are at most of order $||G||$. And, since this norm is finite (i.e. $O(L^0)$), we don't have to worry about the large values of $\lambda_m(\omega)$ arising due to the unbounded tan function. From the form of $P(\lambda)$, we thus conclude that the relevant level spacing is $\Delta\lambda \approx \Delta E \sim \frac{1}{\mathcal{D}}$, where ΔE is the level spacing associated with H_0. The criterion for the localization of the eigenstates $|\chi_i\rangle$ in the energy space of H_0 thus takes the form:

$$|G_{mn}| = |\langle E_m|G|E_n\rangle| \ll 1/\mathcal{D}. \tag{2.15}$$

And, using Eq. 2.8, we find that the Floquet eigenstates $|\psi_i\rangle$ are localized (delocalized) if $|\chi_i\rangle$ are localized (delocalized). We now explore this criterion for delocalization when $\hat{H}_0$ describes an ergodic or a many-body localized phase, since the off-diagonal elements behave quite differently in the two cases.

2.4.1 Ergodic systems

The condition determined by Eq. 2.15 does not hold for ergodic systems even for arbitrarily weak driving. To see why, note that eigenstates $|E_m\rangle$ in the ergodic phase are like random vectors in the basis of product states (except for the energy constraint) according to the ETH. The typical matrix element of *any* local operator between two random vectors is:

$$|G_{mn}| \sim ||G||/\sqrt{\mathcal{D}} \gg 1/\mathcal{D}. \tag{2.16}$$

Thus, in the thermodynamic limit, the hopping problem is always in the delocalized phase.

In the delocalized phase, the Floquet eigenstates are superpositions of nearly all eigenstates $|E_m\rangle$ with macroscopically different physical energies E_m (but with $|\lambda_m(\omega)| \lesssim ||G||$).

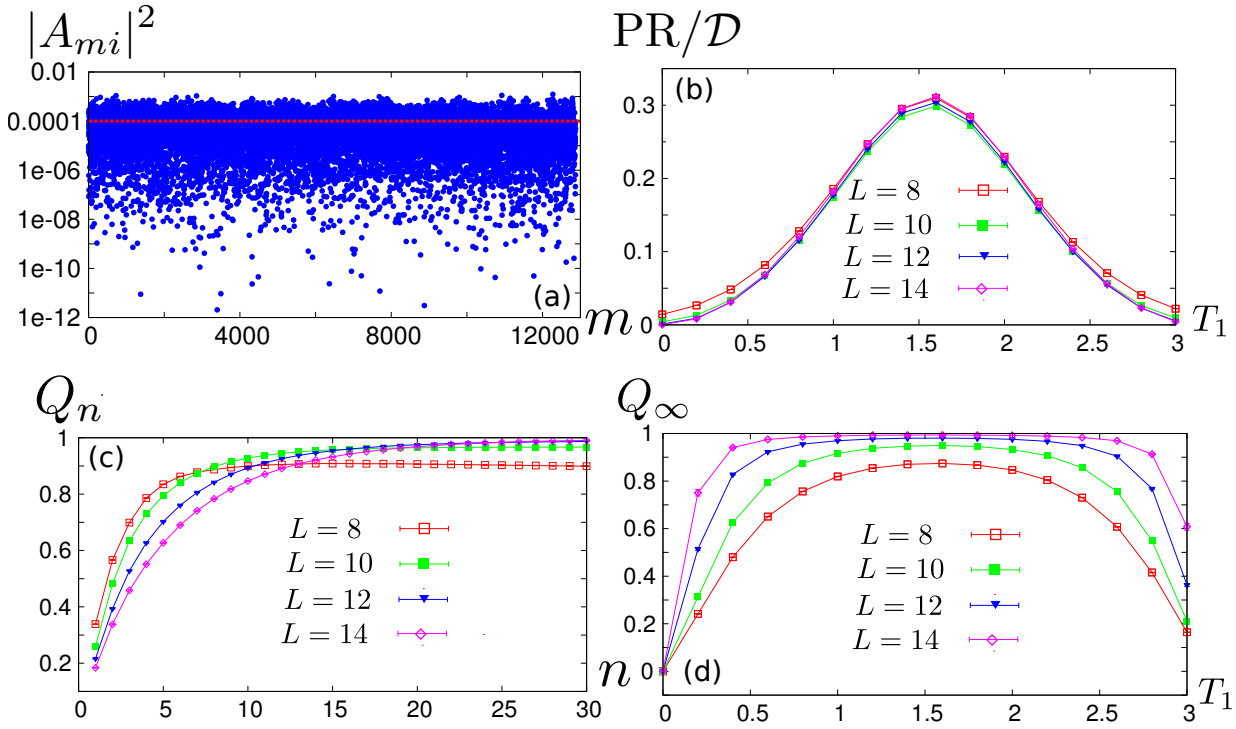

Figure 2.3: Energy delocalization and heating to infinite temperature in the ergodic phase of the XXZ model ($T_0 = 7, W = 0.5$). (a) Squared overlap $|A_{mi}|^2$ of the Floquet eigenstate $|\psi_i\rangle$ with eigenstates $|E_m\rangle$ of H_0, ordered by energy, for a fixed disorder realization and $T_1 = 1.5$. $|A_{mi}|^2$ are nearly uniformly spread over all eigenstates $|E_m\rangle$. (b) Disorder averaged PR/$\mathcal{D}$ vs T_1. As L is increased, PR remains finite in the ergodic phase. (c) Disorder averaged Q_n vs the number of driving cycles n ($T_1 = 1.5$), for evolution starting from the ground state. (d) Disorder averaged saturation value Q_∞ vs T_1 for different L. In the ergodic case, Q_∞ sharply approaches 1 as $L \to \infty$ for any T_1, signaling generic heating to infinite temperature. Number of disorder averages is 2×10^4 ($L = 8, 10$) and $\sim 10^3$ ($L = 12, 14$).

Their PR, defined in Eq. 2.12, remains a finite fraction of the total size of the Hilbert space as $\mathcal{D} \to \infty$, thus individual Floquet eigenstates describe infinite temperature states of the system.

We have verified the above statements numerically in the case of the XXZ model defined in Eq. 2.13. In Fig. 2.3(a) we first illustrate the structure of a typical Floquet eigenstate $|\psi_i\rangle$ for a fixed disorder realization and $T_1 = 1.5$. We plot $|A_{mi}|^2$ as function of $|E_m\rangle$,

ordered by energy E_m. As expected for the ergodic case, a typical Floquet eigenstate is delocalized and has non-zero overlap with states at very different energies.

The difference between Floquet eigenstates is further revealed in the behaviour of PR shown in Fig. 2.3(b). The disorder averaged PR/$\mathcal{D}$, plotted as a function of T_1 for different system sizes L, shows that Floquet eigenstates occupy a finite fraction of the Hilbert space in the thermodynamic limit when the system is ergodic.

The energy absorbed after n cycles, Q_n, defined in Eq. 2.11, when the system is initially prepared in the ground state of H_0, is shown in Fig. 2.3(c). In the thermodynamic limit, Q_n approaches 1 for ergodic systems [1].

Similarly, the saturated value Q_∞ in the ergodic phase (Fig. 2.3(d)) tends to 1 as L is increased. Note that for the system sizes studied here, $Q_\infty \ll 1$ for $T_1 \to 0$ and $T_1 \to \pi$. For a finite size system, at small T_1, the off-diagonal elements are larger than the level spacing only for $\tan(T_1/2) \gg 1/\sqrt{\mathcal{D}}$ which marks the onset of delocalization. As $T_1 \to \pi$, the off-diagonal elements increase as $\tan(T_1/2)$ and the level spacing of the relevant λ_m grows as well, see discussion in the third paragraph of Section 2.4. Assuming delocalization, the relevant λ_α are of order $\bar{\lambda} \sim \tan(T_1/2)\, 1/\sqrt{\mathcal{D}}$ and the level spacing at $\bar{\lambda}$ is $\Delta\lambda \sim (1+\bar{\lambda}^2)\frac{T_0}{2}\frac{1}{\mathcal{D}}$. Further assuming $\bar{\lambda}^2 \gg 1$, i.e. $\tan(T_1/2) \gg \sqrt{\mathcal{D}}$, the delocalization condition reads $\tan(T_1/2)\,\frac{1}{\sqrt{\mathcal{D}}} \gg \Delta\lambda$. This implies that delocalization is not possible for $\tan(T_1/2) \gtrsim \frac{2}{T_0}\mathcal{D}^{3/2}$, which explains the numerical observations. However, Q_∞ monotonically increases as a function of L, and is likely to reach 1 in the thermodynamic limit for arbitrary values of T_1, except for specific values where e^{-iVT_1} is the identity, suggesting that ergodic systems generally heat up to infinite temperature.

2.4.2 Many-body localized systems

The situation for MBL states is quite different. A given MBL eigenstate of $\hat{H}_0$, $|E_m\rangle$ with energy E_m is labelled by the set of values of the l-bits $|\tau\rangle$, where we use the notation $\tau = (\tau_1^z, \ldots, \tau_i^z, \ldots, \tau_L^z)$ (c.f. Eq. 1.20). A local operator G, for e.g. with support only on a specific site i, can be expanded in the basis of Pauli matrices associated with the l-bits: $G = \sum_{j,\alpha} K_{ij}^{\alpha}\tau_j^{\alpha} + \sum_{jk,\alpha\beta} K_{ijk}^{\alpha\beta}\tau_j^{\alpha}\tau_k^{\beta} + \ldots$. The coefficients $K_{i\{j\}}^{\{\alpha\}}$, where $\{\alpha\}$ denotes the list $\alpha_1, \alpha_2, \ldots$ and $\{j\}$ the list $j_1, j_2, \ldots$ of sites, decay exponentially as $e^{-R/\xi}$ with the

[1] We note that, since we consider finite systems, the system will undergo rare quantum revivals when its energy becomes close to the initial value, which is much lower than the infinite temperature value ($Q_n < 1$). However, the revival time increases exponentially with system size and therefore for all practical purposes revivals can be ignored.

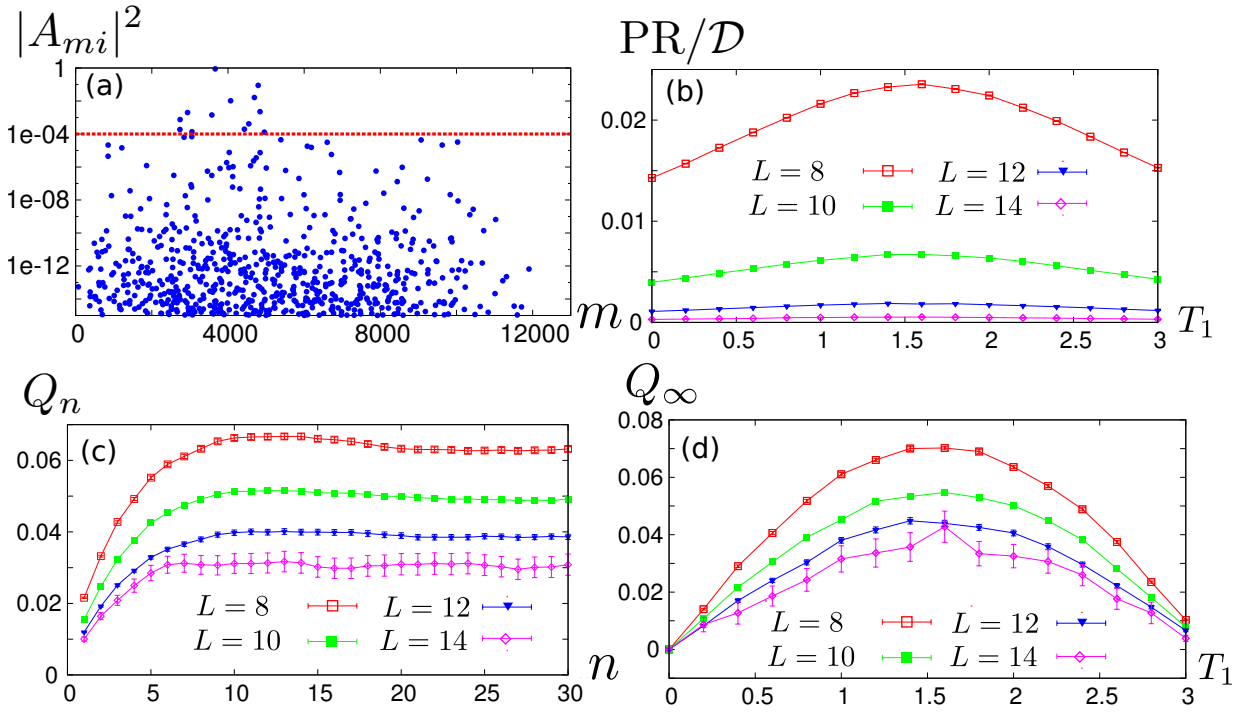

Figure 2.4: Energy localization and absence of heating in the MBL phase ($T_0 = 7, W = 8$). (a) Squared overlap $|A_{mi}|^2$ of the Floquet eigenstate $|\psi_i\rangle$ with eigenstates $|E_m\rangle$ of H_0, ordered by energy, for a fixed disorder realization and $T_1 = 1.5$. The overlap is non-zero only for a few eigenstates with similar energies. (b) Disorder averaged PR/$\mathcal{D}$ vs T_1. As L is increased, PR/$\mathcal{D}$ decreases as $1/\mathcal{D}$. (c) Disorder averaged Q_n vs. the number of driving cycles n ($T_1 = 1.5$), for evolution starting from the ground state. (d) Disorder averaged saturation value Q_∞ vs. T_1 for different L. In the MBL phase, $Q_\infty \ll 1$ and decreases with L, indicating that the system absorbs finite energy locally. Number of disorder averages is 2×10^4 ($L = 8, 10$) and $\sim 10^3$ ($L = 12, 14$).

maximum of distance R between the sites in i and $\{j\}$ on the scale ξ defining a many-body localization length.

Thus, let's consider the states $|\tau'\rangle$ which differ from the fixed state $|\tau\rangle$ only in integrals of motion within a distance R from the support of operator G. The density of these states is dominated by those that differ by $k = R/2$ integrals of motion from $|\tau\rangle$, due to the properties of the binomial coefficient $\binom{R}{k} \sim \binom{R}{R/2} e^{2(R/2-k)^2/R} \sim \frac{2^R}{\sqrt{R}} e^{2(k-R/2)^2/R}$. Hence, the level spacing of these states at energies $E_m + \frac{2\pi}{T_0}$ scales as $\Delta E \sim \frac{WR}{2^R}$, since the bandwidth

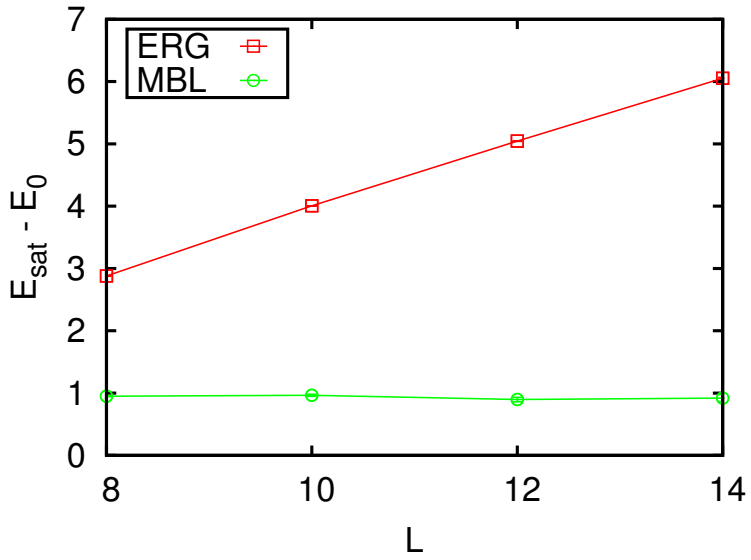

Figure 2.5: Absorbed energy $E_{\rm sat} - E_0 = {\rm Tr}(\rho_\infty H_0) - E_0$ as a function of system size L for ergodic (red) and MBL (green) phases; E_0 is the energy at time $t = 0$. In the former case, absorbed energy is extensive and scales linearly with L, while in the latter case energy is absorbed only locally and its value is independent of L.

scales as $\sim W\sqrt{R}$ due to the random disorder and assuming $W\sqrt{R} \gg \frac{2\pi}{T_0}$. On the other hand, the matrix element with $|\tau\rangle$ is typically $||G||e^{-R/2\xi}$. Therefore, at sufficiently strong disorder and correspondingly large ξ, the level spacing is asymptotically larger in R than the matrix element.

Thus, the criterion for localization (defined in Eq. 2.15) is satisfied in the MBL phase. Hopping G only significantly mixes a few eigenstates with a similar structure away from the support of G. This implies that energy can be absorbed only in the vicinity of the driving.

The structure of the Floquet eigenstates in the MBL phase is thus very different from the ergodic case. A typical Floquet eigenstate in the MBL case has sizable overlap only with those $|E_m\rangle$ that are close in energy, as shown in Fig. 2.4(a). In this case, we fix the disorder realization, and set $T_1 = 1.5$, $T_0 = 7$, and disorder strength $W = 8$. Moreover, the disorder averaged PR/$\mathcal{D}$ approaches zero as $1/\mathcal{D}$ (Fig. 2.4(b)).

The energy absorbed after n cycles, Q_n, when the system is initially prepared in the ground state of H_0, is shown in Fig. 2.4(c). As expected for the MBL case, Q_n is much

smaller than 1 for all T_1 and decreases with system size. Q_∞ remains smaller than one for all T_1, and decays as $1/L$ (Fig. 2.4(d)). These features reflect the local absorption of energy in the system. The finite size scaling of the absorbed energy for the ergodic and MBL cases is shown in Fig. 2.5.

Finally, we have also numerically studied the level statistics of the sorted quasi-energies ω_i of the Floquet operator, characterized by the parameter [30]

$$r_i = \frac{\min(\Delta\omega_i, \Delta\omega_{i+1})}{\max(\Delta\omega_i, \Delta\omega_{i+1})}, \tag{2.17}$$

where $\Delta\omega_i = \omega_i - \omega_{i-1}$. In the ergodic phase, the average r is approximately 0.53, reflecting the Circular Orthogonal Ensemble [58], while in the MBL case, the average r is approximately 0.386, consistent with Poisson statistics. The circular ensembles generalize RMT to the Floquet setting. For the system considered here, the distribution of r is specifically described by the Circular Orthogonal Ensemble due to an anti-unitary symmetry which squares to one and leaves the Floquet operator invariant [59].

2.5 Conclusions

We have shown that periodic local driving has very different effects on ergodic versus MBL systems. Driven ergodic systems heat up to the infinite temperature and their Floquet eigenstates are delocalized in energy space, while MBL systems absorb energy only locally and provide an example of dynamical localization.

The above results indicate that driven interacting systems differ significantly from driven non-interacting systems (on a lattice): the latter have a set of conserved quantities, a local Floquet Hamiltonian and their long time behaviour is described by the Periodic Gibbs Ensemble (c.f. Section 1.4.3) rather than an infinite temperature state [60].

A few remarks are in order. First, we expect that our results hold for other choices of the local operator V and Hamiltonian H_0 in addition to the type of driving protocol. In particular, our results are expected to hold for harmonic driving as well, e.g. $H(t) = H_0 + V\cos(\omega t)$. Second, our approach can be extended to the case of global driving – that is, when V is a sum of local terms. In this case, G is no longer bounded as $L \to \infty$, which can only help with delocalization [58]. Thus globally driven ergodic systems are also expected to heat up to the infinite temperature and to have delocalized Floquet eigenstates in agreement with a recent study [61]. Moreover, our results for ergodic systems suggest that non-interacting topological Floquet bands will generally be unstable to the inclusion of interactions. The case of globally driven MBL systems however is more intricate and deserves a separate study (see Chapter 3).

Chapter 3

Periodically-driven ergodic and many-body localized quantum systems: global drive

3.1 Introduction

This chapter extends the work presented in the previous chapter concerning periodically driven systems under local drive to the setting of global drive.

In this setting, we consider a generic class of periodically driven one dimensional models with quenched disorder and find that as the system's parameters are varied, two distinct phases are realized, which differ in the structure of their Floquet eigenstates as well as in their dynamical properties. One of them is the MBL phase in which the Floquet eigenstates at arbitrary quasi-energy obey the area law for entanglement entropy and thus in this respect behave as ground states of gapped systems. Level repulsion is absent and the statistics of quasi-energy levels follows the Poisson distribution. Further, in the limit of an infinite system, eigenstates with similar quasi-energies typically have different local properties and thus the ETH breaks down. The second phase is the delocalized (ergodic) phase. Here the Floquet eigenstates have volume law entanglement, the quasi-energy levels repel and their statistics are described by the Circular Orthogonal Ensemble (COE). ETH holds in this phase and consequently the Floquet eigenstates have identical local properties and are described by an infinite temperature Gibbs ensemble.

The two phases can furthermore be distinguished by their dynamical properties when the system is initially prepared for example in a product state. In the MBL phase, the

time evolved states retain local memory of the initial state and at long times local observables are correlated with their initial values. As in MBL systems with time independent Hamiltonians, entanglement entropy grows logarithmically in time. This behaviour reflects the presence of emergent local integrals of motion (c.f. Section 1.2), which we explicitly construct following Ref. [69] (see also Ref. [70]). In contrast, in the delocalized phase, local observables relax to their "equilibrium" values at long times, which are given by the infinite temperature Gibbs ensemble. In this case, the entanglement spreading is much faster and consistent with linear growth.

The organization of this chapter is as follows. In the next section (Section 3.2), we introduce the model describing a 1/2-spin chain with open boundary conditions as well as the driving protocol which realizes the two phases mentioned above. In Section 3.3, we probe different properties of the Floquet quasi-energies and eigenstates as the strength of the kick of the drive is varied; these include the gap statistic of the quasi-energies, the half-chain entanglement entropy and we test directly if the ETH is verified for a set of local operators. Next, in Section 3.4, we analyze the system's dynamics by considering a quench experiment where the initial state is a Neel product state; in particular, we study the time dependence of a local observable and the growth in time of the half-chain entanglement entropy. In Section 3.5, we construct integrals of motion following [69] and verify their locality in the MBL phase. Finally, we summarize our findings in Section 3.6.

3.2 Model

Our system is a one dimensional 1/2-spin chain with open boundary conditions. We consider a driving protocol wherein the system's Hamiltonian is periodically switched between two operators, H_0 and H_1, both of which are sums of local terms. An example of a disordered Hamiltonian H_0, which describes an MBL phase and acts for time T_0, is

$$H_0 = \sum_i h_i \sigma_i^z + J_z \sigma_i^z \sigma_{i+1}^z, \tag{3.1}$$

where random fields h_i are uniformly distributed in the interval $[-W, W]$. The eigenstates of H_0 are product states. As a delocalizing Hamiltonian H_1 we choose

$$H_1 = J_x \sum_i \sigma_i^x \sigma_{i+1}^x + \sigma_i^y \sigma_{i+1}^y, \tag{3.2}$$

which acts for time T_1 such that the driving period is $T = T_0 + T_1$. The associated Floquet operator is given by:

$$F = e^{-iH_0 T_0} e^{-iH_1 T_1}. \tag{3.3}$$

The protocol describes an MBL system periodically "kicked" with a delocalizing perturbation H_1, and can be viewed as a many-body generalization of a periodically kicked rotor model (c.f. Section 1.4.1). Recent work has argued that a similar protocol for translationally invariant Hamiltonians results in an infinite temperature state at long times, and therefore a non-local Floquet operator [58]. We fix $J_x = J_z = 1/4$, $T_0 = 1$, $W = 2.5$ and tune the strength of the kick – T_1 – observing a transition at critical T_1^* between an MBL phase (small $T_1 < T_1^*$) and an ergodic phase ($T_1 > T_1^*$). We note that the model defined by Eqs. 3.1 and 3.2 always has one conserved quantity, the z-projection of the total spin, $S_z = \sum_{i=1}^{L} \sigma_i^z$. However, we have checked that this global conservation law is not essential for the existence of the MBL and ergodic phases by studying other models where S_z is not conserved.

3.3 Properties of Floquet eigenstates and quasi-energies

3.3.1 Quasi-energy gap statistics

We first explore the properties of the Floquet eigenstates $|\psi_i\rangle$ and quasi-energy spectrum ω_i associated with the Floquet operator defined in Eq. 3.3, using exact diagonalization (ED). By computing the consecutive quasi-energy gaps $\Delta\omega_i = \omega_i - \omega_{i-1}$, we characterize the level statistics by their ratio $r_i = \min(\Delta\omega_i, \Delta\omega_{i+1})/\max(\Delta\omega_i, \Delta\omega_{i+1})$ (c.f. Section 1.3.1 and Eq. 2.17). The averaged value of r serves as a probe of ergodicity breaking: it allows one to distinguish between the Poisson and Wigner-Dyson level statistics. In Fig. 3.1 we show $\langle r \rangle$ averaged over all quasi-energy spacings and over 1000 disorder realizations, for several system sizes. At small kick period T_1, $\langle r \rangle$ becomes increasingly close to the Poisson value $r_{\text{Poisson}} \approx 0.39$ as the system size is increased. This indicates the absence of level repulsion and suggests that ergodicity is broken at small T_1 and the system is in the MBL phase. On the other hand, at large T_1, $\langle r \rangle$ is approximately equal to 0.53, which is close to the COE value, $r_{\text{COE}} \approx 0.53$ [58]. This suggests that at large T_1 the system delocalizes. The $\langle r \rangle$ curves for different system sizes cross at $T_1^* \approx 0.9$, suggesting a phase transition between MBL and ergodic phases. A drift of the crossing point towards smaller T_1 is observed, similar to the time independent case [30].

3.3.2 Half-chain Entanglement Entropy

To further distinguish the nature of the two phases, we study the entanglement properties of the Floquet eigenstates. The expectation from the static case is that MBL eigenstates

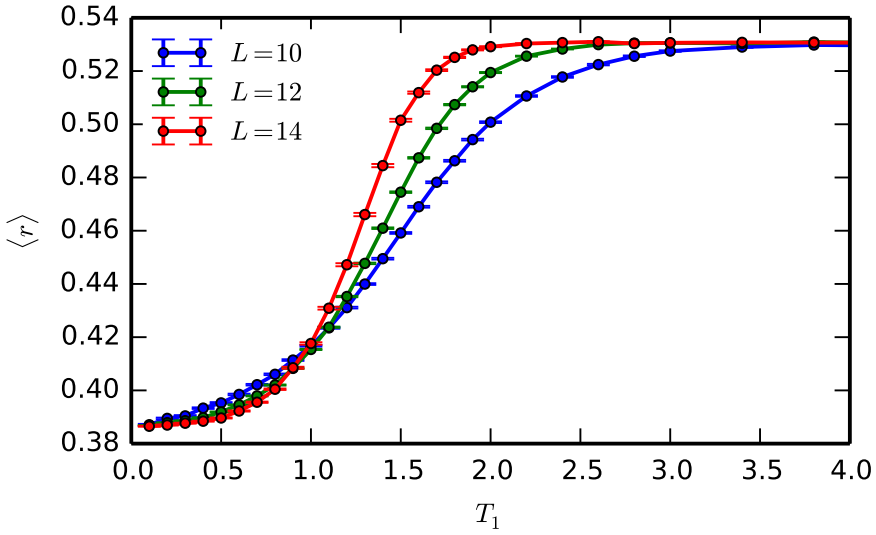

Figure 3.1: Disorder averaged level statistics parameter $\langle r \rangle$ as a function of the "kick" strength T_1. At small values of T_1, $\langle r \rangle \approx 0.386$, indicating Poisson statistics of quasi-energy levels (no level repulsion). At larger T_1 the system undergoes a transition into a delocalized phase with $\langle r \rangle \approx 0.53$, consistent with the COE [58]. The presented data is for system sizes $L = 10, 12, 14$ and averaging is performed over 1000 disorder realizations.

should obey an area law for entanglement entropy, i.e. in 1d their entropy should weakly depend on the chain size [25, 71], while in the ergodic phase the eigenstates are thermal and their entropy is extensive in the subsystem size. Fig. 3.2 shows disorder and eigenstate averaged von Neumann entropy $\langle S \rangle$ of the Floquet eigenstates, for the symmetric bipartition, plotted as a function of T_1. The markedly different scaling of $\langle S \rangle$ at small and large values of T_1 lends further support to the existence of two phases. At $T_1 \lesssim T_1^*$, $\langle S \rangle$ is much smaller than the value expected for random vectors in the Hilbert space, $S_{\rm Th} \approx L/2 \ln 2$ [72], which signals ergodicity breaking. Moreover, at $T_1 \lesssim 0.6$ the entanglement entropy grows very weakly with system size, consistent with area law in 1d. On the contrary, at large $T_1 > T_1^*$, $\langle S \rangle$ approaches $S_{\rm Th}$, indicating that almost all eigenstates are essentially random vectors in the Hilbert space, as expected in the ergodic phase.

It is also instructive to study the fluctuations of entanglement entropy, as they have been shown to provide a useful probe of the MBL-delocalization transition in time independent models [73]. The disorder averaged fluctuations of S, defined as $\Delta S = \sqrt{\langle (S - \langle S \rangle)^2 \rangle}$

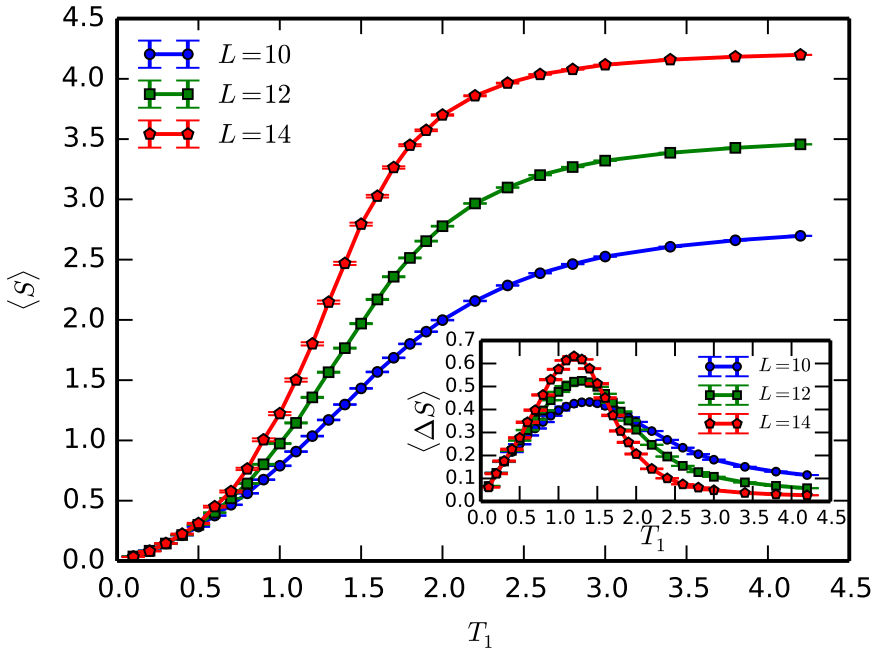

Figure 3.2: Averaged entanglement entropy $\langle S \rangle$ and its fluctuations $\langle \Delta S \rangle$ (inset) as a function of T_1. The scaling of entropy and its fluctuations with system size L are consistent with the existence of an MBL and a delocalized phase for small and large T_1, respectively.

are expected to be small deeply in the delocalized phase, as well as in the MBL phase: in the former case, almost all eigenstates are highly entangled, with $S \approx S_{\mathrm{Th}}$, with small fluctuations around this value, while in the latter case, S obeys an area law and is therefore small, as are its state-to-state fluctuations. In contrast, at the transition S has a broad distribution [25, 73] and therefore its fluctuations are maximal. Thus, the localization-delocalization transition can be detected by the location of the peak in ΔS. Fig. 3.2(inset) shows ΔS as a function of T_1. Entanglement fluctuations ΔS exhibit a maximum at $T_1 \approx 1.1$ that roughly agrees with T_1^* value found from analyzing level statistics; further, we observe a slight drift of the maximum with the system size, similar to the previous study of the static case [73]. We attribute the difference between the position of the maximum in ΔS and value T_1 determined from the level statistics to finite size effects.

3.3.3 Testing the ETH

We have also directly tested the ETH and its violation in the MBL phase in the Floquet eigenstates, finding behaviour consistent with the existence of two phases (see also Ref. [15], where ETH for driven ergodic systems was tested). According to the ETH, in the delocalized phase the expectation value of a local operator $\mathcal{O}$ in all Floquet eigenstates should converge, in the thermodynamic limit, to the prediction of the canonical ensemble with infinite temperature

$$\mathcal{O}_\infty = \frac{1}{\mathcal{D}} \mathrm{Tr}\, \mathcal{O},$$

where $\mathcal{D}$ is the Hilbert space dimension. We test ETH and its violation in the MBL phase by examining the deviation of the expectation value of $\mathcal{O}$ in individual eigenstates, $\langle \mathcal{O} \rangle_i$ from $\mathcal{O}_\infty$:

$$\Delta\mathcal{O} = |\langle \mathcal{O} \rangle_i - \mathcal{O}_\infty|, \tag{3.4}$$

where averaging is performed over all Floquet eigenstates for each disorder realization, and then over different disorder realization. We expect this quantity to approach zero in the delocalized phase, as $L \to \infty$. On the other hand, in the MBL phase this quantity should remain finite as we extrapolate the chain size L to infinity.

In Fig. 3.3 we show numerical results for local operators $\mathcal{O}$ that act on the two neighbouring sites in the middle of the chain and conserve S^z. We studied the following four operators:

$$\begin{aligned}
\mathcal{O}_1 &= \sigma^z_{L/2}, \\
\mathcal{O}_2 &= \sigma^z_{L/2}\sigma^z_{L/2+1}, \\
\mathcal{O}_3 &= (\sigma^+_{L/2}\sigma^-_{L/2+1} + \sigma^-_{L/2}\sigma^+_{L/2+1}), \\
\mathcal{O}_4 &= i(\sigma^+_{L/2}\sigma^-_{L/2+1} - \sigma^-_{L/2}\sigma^+_{L/2+1}).
\end{aligned}$$

As expected, in the MBL phase ($T_1 = 0.4$), $\langle \Delta\mathcal{O} \rangle$ changes weakly with system size suggesting that it remains finite in the thermodynamic limit. In the delocalized phase ($T_1 = 3.0$), $\Delta\mathcal{O}$ approaches zero with increasing system size suggesting that each Floquet eigenstate behaves as an infinite temperature thermal state for local observables in the thermodynamic limit. We note that all four operators show nearly identical behaviour. Thus the direct test of the ETH is consistent with the presence of two phases with markedly different properties of their eigenstates.

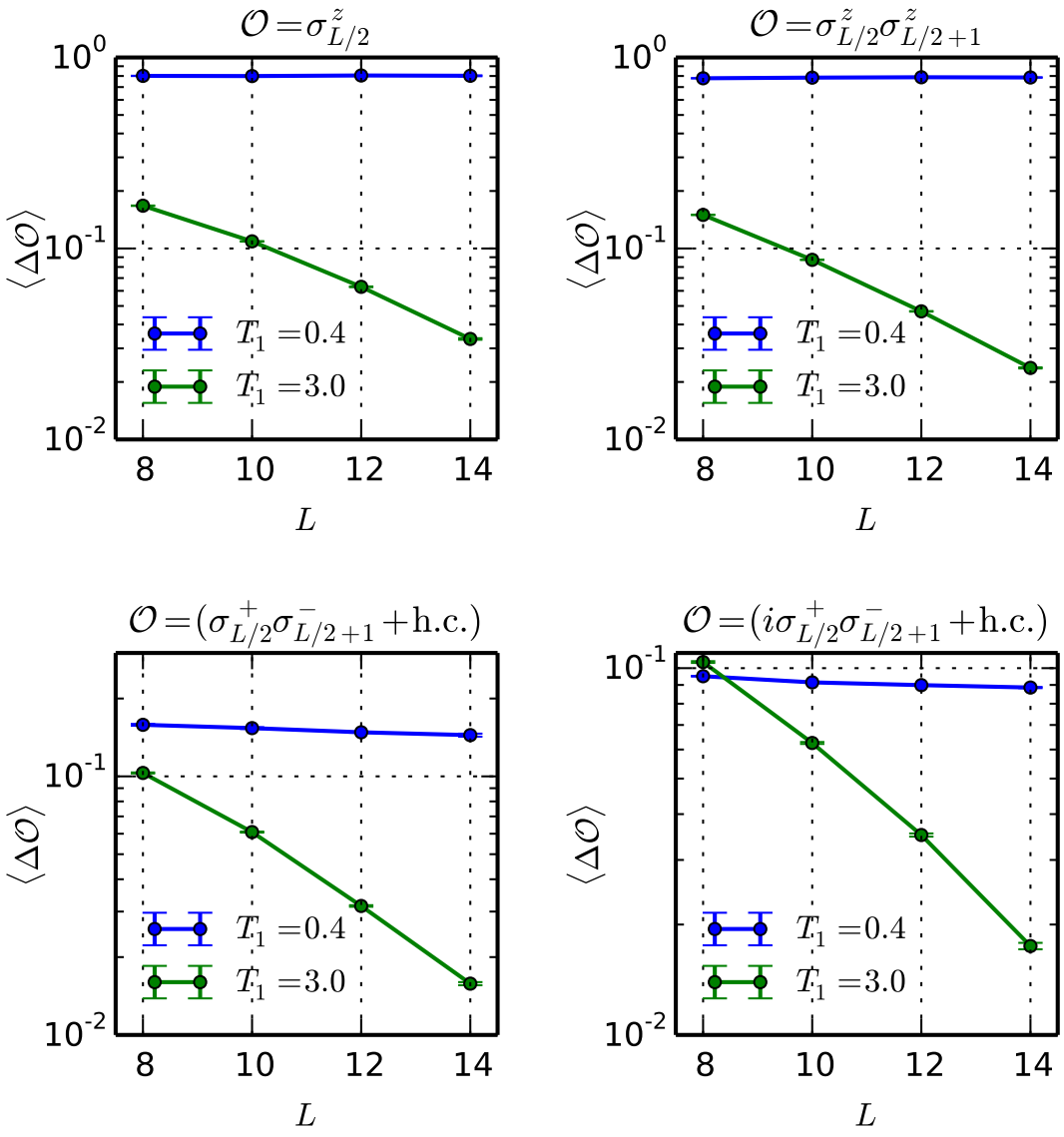

Figure 3.3: Deviation of expectation values of local operators from their infinite temperature values given by the canonical ensemble. Different plots correspond to various choices of S_z-preserving local operators $\mathcal{O}$ acting on the sites in the middle of the chain, and $T_1 = 0.4, 3.0$ correspond to MBL and delocalized phases, respectively.

3.4 Dynamics

We next study the dynamical properties of the model described by Eqs. 3.1 and 3.2. We consider a standard quantum quench protocol: the system is initially prepared in a Néel (product) state $|\varphi_0\rangle$ of spins $\sigma_i^z = \pm 1$ at $t = 0$, and this state is evolved under the time dependent Hamiltonian. This protocol is particularly easy to simulate using Krylov subspace projection methods [74] or the time-evolving block decimation [75] method, both of which allow us to access larger systems beyond ED due to the sufficiently slow growth of entanglement in the MBL phase. For the TEBD algorithm we use a second order Trotter decomposition with time step $\Delta t = 0.1$. The growth of the bond dimension is controlled by requiring the neglected weight to be less than 10^{-7} at each Schmidt decomposition.

3.4.1 Local observables

We first focus on the evolution of local observables, and compute the expectation value of the spin on a given site I, $\sigma_I^z(t)$, and its long time limit $\langle\sigma_I^z(\infty)\rangle$ [76, 69, 77]. Fig. 3.4 illustrates the time evolution $\sigma_I^z(t)$ for the Néel initial state $|\varphi_0\rangle$ and site $I = 1$, and for system sizes ranging from $L = 10 - 14$ (obtained via ED), $L = 16, 18$ obtained using Krylov subspace projection, and $L = 24, 30$ obtained using TEBD. We find that the on-site magnetization remains finite at very long times even for the largest systems without any visible finite size effects. This indicates that the MBL phase remains stable in the thermodynamic limit.

The infinite time average of the operator $\sigma_I^z(t)$ with respect to the initial state $|\varphi_0\rangle$ reads $\langle\sigma_I^z(\infty)\rangle = \lim_{n\to\infty} \frac{1}{n+1} \sum_{k=0}^{n} \langle\varphi_0|\sigma_I^z(kT)|\varphi_0\rangle$, which in terms of the Floquet eigenstates $|\psi_i\rangle$ is given by the expectation value in the diagonal ensemble (c.f. Eq. 1.6), i.e. $\sum_i \langle\psi_i|\sigma_I^z|\psi_i\rangle|\langle\varphi_0|\psi_i\rangle|^2$. The infinite time value $\langle\sigma_I^z(\infty)\rangle$, calculated using ED, and averaged over 6000 disorder realizations, is illustrated in Fig. 3.4(inset). This quantity behaves differently in the two phases: at $T \lesssim T_1^*$, $\langle\sigma_I^z(\infty)\rangle$ is positive and weakly dependent on the system size, which shows that in the MBL phase the local memory of the initial state is retained. Deep in the ergodic phase at $T_1 \gg T_1^*$, however, $\langle\sigma_I^z(\infty)\rangle \to 0$, reflecting the decay of the initial magnetization and therefore a loss of the memory of the initial state.

3.4.2 Entanglement growth

Finally, we explored the spreading of entanglement following a quantum quench, known to be a sensitive probe of many-body localization: in the MBL phase, entanglement grows

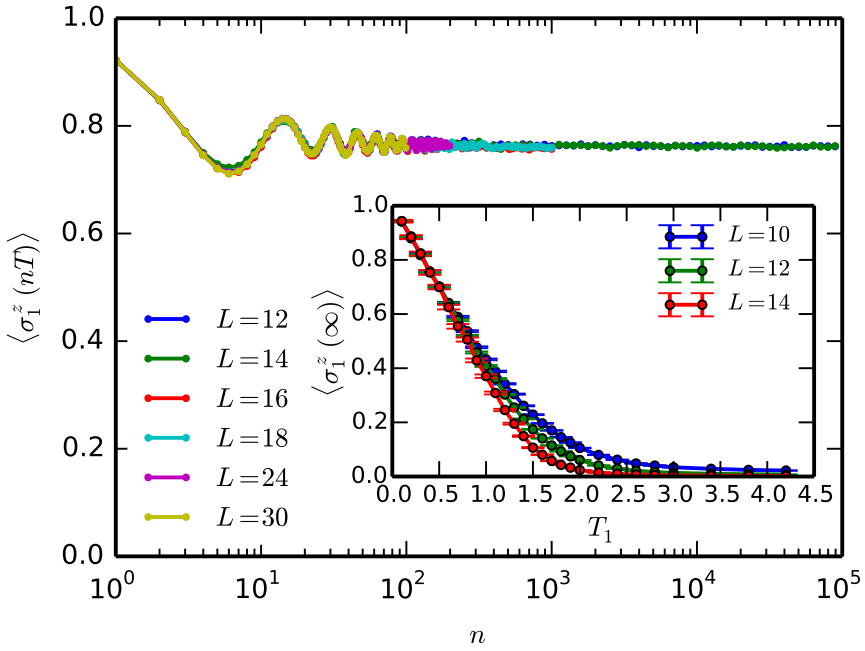

Figure 3.4: Dynamical properties: decay of magnetization at a given site $I = 1$ for a Néel initial configuration. Inset: The long time magnetization remains non-zero in the MBL phase as the system size is increased, while in the delocalized phase it decays to zero. Averaging was performed over 6000 disorder realizations.

logarithmically in time [27, 78, 25, 26, 79], while in the ergodic phase, as well as in Bethe-ansatz-integrable systems, it grows linearly in time [80, 81, 28]. The disorder averaged entanglement entropy as a function of time, calculated for fixed $T_1 = 0.4$ and for the symmetric bipartition, is shown in Fig. 3.5. Averaging was performed over 6000 disorder realizations. Entanglement initially rises from zero, followed by a plateau and a logarithmic growth for several decades in time, $\langle S(t) \rangle \propto \ln(t)$. This behavior is qualitatively similar to that found in the MBL phase in systems with time independent Hamiltonians [78, 25, 27, 26], which gives further support for the existence of the MBL phase in driven systems with strong disorder.

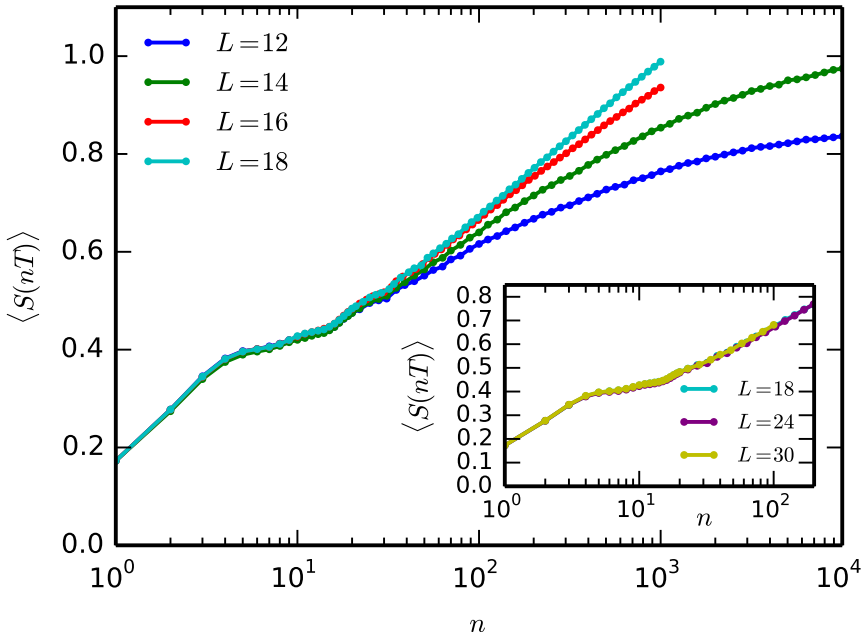

Figure 3.5: Disorder averaged entanglement entropy following a quantum quench, for the Néel initial state. Data for system sizes $L = 12, 14$ was obtained by ED, for $L = 16, 18$ using Krylov subspace projection, and $L = 24, 30$ using TEBD. Averaging performed over 6000 disorder realizations.

3.5 Local integrals of motion and effective description of the driven MBL phase

In order to understand the spectral and dynamical properties of the MBL phase observed in the numerical simulations, we propose that this phase is characterized by an extensive number of local integrals of motion [25, 26]. First, we note that the area law entanglement of the Floquet eigenstates suggests that they can be obtained from the product states (in the $\sigma_i^z = \pm 1$ basis) by a quasi-local unitary transformation U which brings the Floquet operator into a diagonal form in that basis: $UFU^\dagger = F_{\text{diag}}$. Since L of the operators σ_i^z commute with F_{diag}, we can introduce a set of L "pseudospin" operators $\tau_i^z = U^\dagger \sigma_i^z U$. These operators commute with the Floquet operator $[F, \tau_i^z] = 0$, as well as with each other $[\tau_i^z, \tau_j^z] = 0$. Operators τ_i^z have eigenvalues ± 1 and therefore satisfy the relation $(\tau_i^z)^2 = 1$; they can be viewed as z-components of some "effective" spins. We emphasize that the

operators τ_i^z can be introduced for any driven system, but the special property of the MBL phase is that their support is localized near site i, and they affect remote physical degrees of freedom exponentially weakly. In terms of τ operators or l-bits, the operator F takes a simple form, as it can only depend on the τ_i^z operators and their products (but not on the τ_i^x, τ_i^y operators). It is convenient to represent F as

$$F = e^{-iH_{\text{eff}}(\{\tau_i^z\})}, \tag{3.5}$$

where $H_{\text{eff}}(\{\tau_i^z\})$ is a real function of operators τ_i^z. (Such a representation takes into account the fact that eigenvalues of F have absolute value one). Further, since $(\tau_i^z)^2 = 1$, H_{eff} can generally be written as

$$H_{\text{eff}}(\{\tau_i^z\}) = \sum_i J_i^{(1)} \tau_i^z + \sum_{ij} J_{ij}^{(2)} \tau_i^z \tau_j^z + \sum_{ijk} J_{ijk}^{(3)} \tau_i^z \tau_j^z \tau_k^z + \dots \tag{3.6}$$

It is natural to assume that in the MBL phase the couplings J between remote effective spins decay exponentially with distance, in analogy to the static case [25, 26]; we note that long-range interactions, in particular, would be inconsistent with Lieb-Robinson bounds on information propagation [82] satisfied by the operator F, which bound the spread of information by a linear function of time.

The effective model introduced above naturally explains the spectral and dynamical properties of the MBL phase established numerically, for example, the absence of decay of the on-site magnetization at long times and the logarithmic growth of entanglement, which directly follow from Eqs. 3.5 and 3.6. The latter observation in particular is due to the exponential decay of interactions between remote effective l-bits [78, 25, 26].

To further support the effective model for the MBL phase, we now explicitly construct an extensive set of quasi-local integrals of motion as in Ref. [69]. Even though these form an extensive set, they are not identical to the τ_i^z operators. In particular, they do not commute with each other, which is a defining property of the latter. Following Ref. [69], we consider the infinite time average of the operator σ_1^z in the Heisenberg picture, denoted by $\bar{\sigma}_1^z$, which is always an integral of motion by definition. To see this, notice that σ_1^z (or for that matter any operator) can be written in the basis of eigenstates of the Floquet Hamiltonian H_F:

$$\sigma_1^z = \sum_{i,j} \langle \psi_i | \sigma_1^z | \psi_j \rangle |\psi_i\rangle\langle\psi_j| \tag{3.7}$$

In this basis, the stroboscopic time evolution after n periods of the operator $\sigma_1^z(nT) \equiv e^{iH_F nT} \sigma_1^z e^{-iH_F nT}$ in the Heisenberg picture is simply:

$$\sigma_1^z(nT) = \sum_{i,i} \langle \psi_j | \sigma_1^z | \psi_i \rangle e^{-i(\omega_j - \omega_i)nT} |\psi_j\rangle\langle\psi_i| \tag{3.8}$$

The infinite time average defined as $\bar{\sigma}_1^z = \lim_{n\to\infty} \frac{1}{n+1} \sum_{k=0}^{n} \sigma_1^z(nT)$ is given by $\bar{\sigma}_1^z = \sum_i \langle\psi_i|\sigma_1^z|\psi_i\rangle|\psi_i\rangle\langle\psi_i|$; the oscillating factor in the off-diagonal elements averages to zero. This result assumes there are no degeneracies in the phases $\omega_i T$. We now demonstrate that in the MBL phase this integral of motion is a *quasi-local* operator.

First, we note that operator $\bar{\sigma}_1^z$ describes the spreading of magnetization, initially prepared on site 1. To clarify this statement, consider the infinite temperature ensemble in which spin 1 was initially prepared in the up state, described by the density matrix

$$\rho(0) = 2^{-L}(1 + \sigma_1^z) \otimes 1_{i\neq 1}. \tag{3.9}$$

Note that $\rho(0)$ is a maximum entropy mixture of the states of the form $|s_1 = 1, \{s\}\rangle$, where $\{s\}$ denotes the arbitrary value of σ_i^z for spins $2, \ldots, L$. The total magnetization $\sum_{i=1}^{L} \sigma_i^z$ of this state is 1, since $\mathrm{Tr}\rho(0)\sigma_1^z = 1$ and $\mathrm{Tr}\rho(0)\sigma_i^z = 0$ for $i \neq 1$.

As the dynamics conserves the total magnetization, the initial excess magnetization located on spin 1 will spread over the rest of the chain with time. Upon time averaging, the density matrix can be expressed in terms of the operator $\bar{\sigma}_1^z$ as follows:

$$\bar{\rho} = 2^{-L}(1 + \bar{\sigma}_1^z). \tag{3.10}$$

Thus, the spreading of the initial magnetization over the chain at long times can be related to the properties of the operator $\bar{\sigma}_1^z$; the long time magnetization on site j is given by:

$$M_{1j} = \mathrm{Tr}(\bar{\rho}\sigma_j^z) = \frac{1}{2^L}\mathrm{Tr}\left(\bar{\sigma}_1^z\sigma_j^z\right). \tag{3.11}$$

Fig. 3.6 illustrates that M_{11} is on the order of unity, but M_{1j} decays over several orders of magnitude as a function of $|j-1|$ in the MBL phase ($T_1 = 0.4$). This is consistent with $\bar{\sigma}_1^z$ being a quasi-local operator. Conversely, in the delocalized regime ($T_1 = 3$), the magnetization is nearly uniformly spread over all sites j and has a stronger dependence on the chain size L, thus, the operator $\bar{\sigma}_1^z$ is non-local.

To further test the quasi-locality of operator $\bar{\sigma}_1^z$ in the MBL phase, we examined the partial norm

$$\mathcal{N}(j) = \frac{1}{2^j}\mathrm{Tr}(\bar{\sigma}^A\bar{\sigma}^A), \tag{3.12}$$

where

$$\bar{\sigma}^A \equiv \frac{1}{2^{L-j}}\mathrm{Tr}_{\bar{A}}\bar{\sigma}_1^z, \tag{3.13}$$

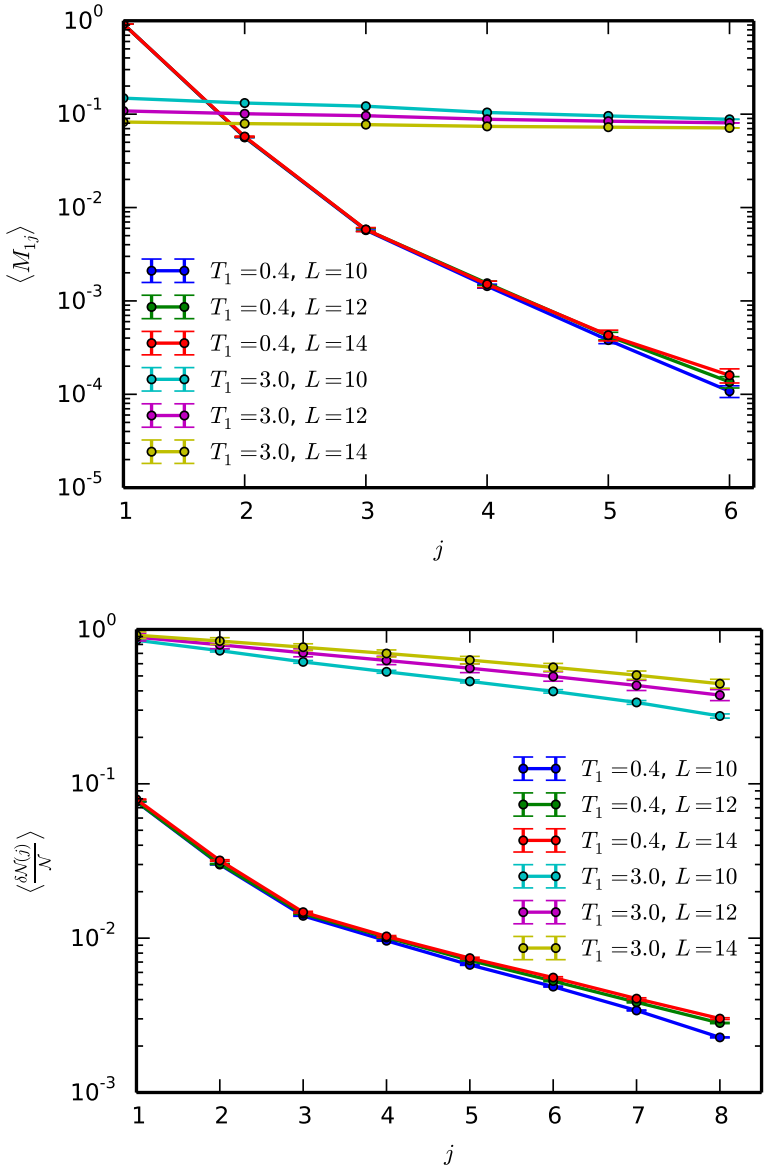

Figure 3.6: Local integrals of motion in the MBL and delocalized phase. (Left) Median magnetization M_{1j} as a function of distance $|1-j|$. (Right) Median difference between the total norm $\mathcal{N}$, and partial norm $\delta\mathcal{N}(j)$, divided by $\mathcal{N}$. The exponential decay of this quantity with distance $|j-1|$ demonstrates that in the MBL phase ($T_1 = 0.4$) operator $\bar{\sigma}_1^z$ is a quasi-local integral of motion. In the delocalized phase ($T_1 = 3$), this operator becomes non-local.

and A is the region containing sites 1 to j and $\bar{A}$ its complement. To understand the physical meaning of $\bar{\sigma}^A$ notice the following property of the partial trace for an operator of the product form $\mathcal{O}_A \otimes \mathcal{O}_{\bar{A}}$: $\mathrm{Tr}_{\bar{A}}\mathcal{O}_A \otimes \mathcal{O}_{\bar{A}} = \mathcal{O}_A \mathrm{Tr}_{\bar{A}}\mathcal{O}_{\bar{A}}$, which is shown below:

$$\langle\{s'_A\}|\mathrm{Tr}_{\bar{A}}\mathcal{O}_A \otimes \mathcal{O}_{\bar{A}}|\{s_A\}\rangle = \sum_{\{s_{\bar{A}}\}} \langle\{s'_A\}\{s_{\bar{A}}\}|\mathcal{O}_A \otimes \mathcal{O}_{\bar{A}}|\{s_A\}\{s_{\bar{A}}\}\rangle \tag{3.14}$$

$$= \langle\{s'_A\}|\mathcal{O}_A|\{s_A\}\rangle \sum_{\{s_{\bar{A}}\}} \langle\{s_{\bar{A}}\}|\mathcal{O}_{\bar{A}}|\{s_{\bar{A}}\}\rangle. \tag{3.15}$$

Hence, it follows that $\bar{\sigma}^A$ is a truncation of the operator $\bar{\sigma}^z_1$ to the region A in the sense that it is obtained from the latter by removing the terms of the expansion of $\bar{\sigma}^z_1$ in Pauli strings $\sigma^{\alpha_1}_{i_1} \dots \sigma^{\alpha_k}_{i_k}$ with $k = 1, 2, \dots, L$ that do not act as the identity outside of A, and keeping the remaining terms as is.

In Fig. 3.6, we illustrate the normalized difference $\frac{\delta\mathcal{N}(j)}{\mathcal{N}} = \frac{\mathcal{N}-\mathcal{N}(j)}{\mathcal{N}}$ vs j, where $\mathcal{N} = \frac{1}{2^L}\mathrm{Tr}\bar{\sigma}^z_1\bar{\sigma}^z_1$ is the total norm of the operator $\bar{\sigma}^z_1$. This quantity describes how well the operator $\bar{\sigma}^z_1$ can be approximated by operators with a finite support, and therefore tests whether this operator is quasi-local. It is evident from Fig. 3.6 that $\frac{\delta\mathcal{N}(j)}{\mathcal{N}}$ approaches zero exponentially in distance $|j-1|$ in the MBL phase, indicating that the operator $\bar{\sigma}^z_1$ is indeed a quasi-local integral of motion. We note that similar quasi-local integrals of motion, $\bar{\sigma}^z_i$, can be constructed for other sites, $i = 2, .., L$, and they form an extensive set of LIOMs.

3.6 Conclusions

We have demonstrated the existence of two dynamical regimes in periodically driven systems described by local interacting Hamiltonians with quenched disorder. In particular, we have identified a many-body localized phase in which ergodicity is broken. We argued that the MBL phase is characterized by an extensive number of emergent quasi-local conservation laws. This implies that the dynamics of Floquet MBL systems is described by an effective quasi-local time independent Hamiltonian H_{eff}, which is itself many-body localized. This is in sharp contrast to the ergodic phase where the Floquet Hamiltonian does not have a quasi-local representation [58, 61, 1]. An interesting open question is whether the Magnus expansion [53] converges in the MBL phase.

Another implication of our results is that MBL does not rely on global conservation laws such as energy conservation. Further, the MBL phase is robust under sufficiently weak periodic driving and there is a finite driving threshold above which transport is restored and the system eventually delocalizes. This may serve as an experimental signature of many-body localization.

Chapter 4

Thermal inclusions: how one spin can destroy a many-body localized phase

4.1 Introduction

While MBL phases have been essentially proven to exist in one dimensional systems [83], their existence and stability in higher dimensions remains controversial. Indeed, a couple of works [84, 85] have put forward different scenarios whereby MBL phases in higher dimensions are destabilized by small thermal regions. Ref. [84] studied the effect of a thermalizing boundary on a $d > 1$-dimensional MBL phase and argued that the l-bits and other eigenstate measures of localization are unstable; instead, approximately conserved l -bits underlie what remains of localization in such systems. In Ref. [85] the authors argue that any nonzero density of large enough thermalizing inclusions will destroy the MBL phase in $d > 1$, converting it to an extremely slow, but ultimately thermalizing regime.

One of the motivations of the work presented in this chapter is to study the key step in the destabilization process described above in $d > 1$; that is, how a single inclusion interacting with the same strength J with many l-bits on its border leads to their thermalization. We show that a single two-level inclusion is sufficient to thermalize N neighboring l-bits in the large-N limit at any nonzero J/W, where W is the strength of the random fields on the l-bits. In order for such a "central spin" system to remain in the MBL phase in this large-N limit, the interaction with the inclusion must be scaled down with increasing N. Specifically, we find that the critical $J_c(N)$ separating the MBL phase for $J < J_c(N)$ from the thermalizing phase at $J > J_c(N)$ scales as $J_c(N) \sim W/N$ up to multiplicative $\ln(N)$ corrections (see Fig. 4.1).

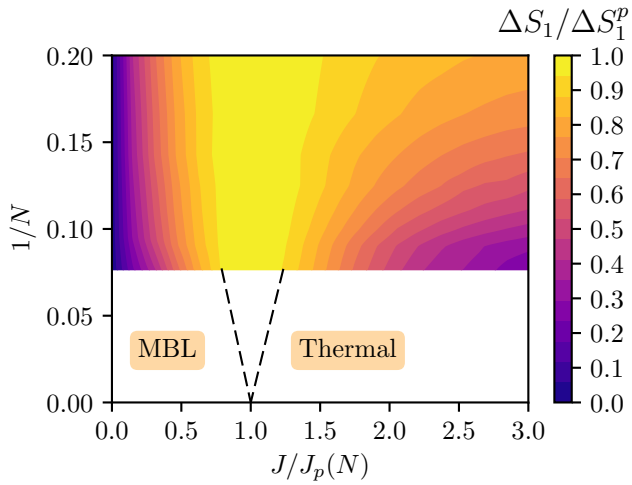

Figure 4.1: Schematic finite size phase diagram of the central spin model mapped out through the standard deviation ΔS_1 over l-bits, eigenstates and samples of the single site entanglement entropy under the simplest assumption of a single transition at $J_c = J_p(N)$.

We note that at the transition for our central spin model, the interactions are crucial to the dynamics of the system but do not contribute to the system's $N \to \infty$ equilibrium thermodynamics, as the number of different interaction terms are $O(N)$ and the coupling J would need to be much larger (of order $O(N^0)$) to do so.

In the sequel, we analyze the infinite temperature dynamical phase diagram of our central spin model using exact diagonalization with up to $N = 13$ l-bits, and using small J perturbation theory at large N. Numerically, the single-site eigenstate entanglement entropy, energy level repulsion and many-body eigenstate participation ratios all support the existence of a MBL phase for $J < J_c(N) \sim W/N$ and an ergodic phase for $J > J_c(N)$. They also reveal many interesting features about the localized phase and the crossover region at $J \approx J_c(N)$. In the localized phase, the mean single-site eigenstate entanglement entropy $[S_1]$ decreases as $1/N$, while the participation ratio distributions are N-independent.

In the crossover region, the l-bits are either "on" (strongly entangled) or "off" (weakly entangled) in eigenstates and the pattern of "on" l-bits varies significantly between states of the same sample (and of course, between samples). Thus, single-site observables are very heterogenous in real space, in energy space and across disorder realizations in the crossover region, suggesting that they change discontinuously as $N \to \infty$, in line with recent proposals [86, 87] that few body observables are similarly discontinuous across the

MBL transition in one dimension.

The perturbative analysis at small J on the classical hypercube explains many of these numerical observations. At second order, a typical initial configuration is resonant with $K \sim J^2N^2/W^2$ states in which two l-bits are flipped and the central spin is flipped. These states are in turn resonant with $\sim K$ *completely new* states at the next order. The resulting 'resonant subgraph' is therefore locally tree-like and we argue that the statistical properties of the resulting eigenstates in the localized phase can be understood via an associated bond percolation problem on the hypercube.

This mapping however does not capture the transition region, in particular, the heterogeneity in single-site observables. This is not particularly surprising as we have neglected higher order processes. Ref. [91] treats these processes within the tree approximation in a related model. Their arguments seem to predict an intermediate delocalized non-ergodic phase between $J_c(N) \sim W/(N \ln N)$ and $J^*(N) \sim W/N$, that is, between the MBL phase and the thermal phase in the central spin model. The numerical data at the accessible system sizes is not conclusive about the existence of this possible intermediate phase or the logarithmic suppression of the critical coupling. We speculate on the possible phase diagrams in the thermodynamic limit in Section 4.5.6.

One difference between our central spin model in the limit of large N and a short-range model in the limit of large d is due to the order of how these two limits are taken. Suppose we are in large but finite dimension d, with a system of size N spins and with coordination number z. If we keep the coordination number z fixed and take the thermodynamic limit of $N \to \infty$, then our sample is infinite. Thus, it has an infinite number of chances to produce a rare thermalizing inclusion composed of a finite number (i.e. O(1)) of spins. To be clear, one process by which a local thermal inclusion can be generated is when the random disorder fields of a contiguous set of n spins is unusually small. In this case, these n spins will be in the ergodic phase and this process occurs with probability $p_n \sim (W_c/W)^n$. And over a volume of size $V \sim 1/p_n$, we typically find one such inclusion. So for any nonzero spin-flip interaction J these inclusions are present at some nonzero density and they destabilize the MBL phase with probability one. In our central spin model, on the other hand, we take z to infinity along with N, and scale J accordingly, which yields a stable MBL phase.

This chapter has the following outline. We first review in more detail the results of Refs [84, 85] in Section 4.2. Next, we describe the model that we studied in Section 4.3 and present the numerical exact diagonalization results in Section 4.4. We then turn to the perturbative analysis at low orders in Section 4.5.2 and Section 4.5.3 and compare their quantitative predictions for the MBL phase and the crossover region to the numerical results in Section 4.5.5. Finally, we discuss the role of higher order processes in Section 4.5.6.

4.2 Review

Ref. [84] studied a system composed by an MBL bulk coupled to an ergodic boundary, wherein the physical spins $\sigma^x_{L,\boldsymbol{i}}$ at the MBL boundary interact locally with the spins $\gamma^x_{\boldsymbol{i}}$ on the thermal boundary: $H_{\text{int}} = g\sum_{\boldsymbol{i}\in\text{edge}} \sigma^x_{L,\boldsymbol{i}}\gamma^x_{\boldsymbol{i}}$.

In the l-bit picture for the MBL bulk, due to their exponentially localized profile, every l-bit interacts with the thermal boundary notwithstanding the exponentially small interactions for the l-bits far away from the boundary. Accordingly, using l-bit operators the interaction Hamiltonian reads $H_{\text{int}} = g\sum_{\boldsymbol{i}\in edge}\sum_{\boldsymbol{j}} C_{\boldsymbol{ij}}\tau^x_{\boldsymbol{j}}\gamma^x_{\boldsymbol{i}}$, where $C_{\boldsymbol{ij}} \sim e^{-R_{\boldsymbol{ij}}/\xi}$ and $R_{\boldsymbol{ij}}$ denotes the distance between an l-bit localized around site $\boldsymbol{j}$ and a physical spin $\boldsymbol{i}$ on the thermal boundary.

The eigenstates of the uncoupled systems have a product form $|E,\tau\rangle$, where τ represents the state of all the l-bits. The interactions terms $\gamma^x\tau^x$ induce l-bits flips in the bulk and mix different thermal eigenstates of the boundary. To probe the stability of the l-bits with respect to the interaction with the thermal boundary, Ref. [84] studied the rate of mixing between a fixed state $|E,\tau\rangle$ and states where l-bit at position $\boldsymbol{m}$ is flipped, i.e. $|E',\tau'_{\boldsymbol{m}}\rangle$. This is given by the Fermi Golden Rule to leading order in g:

$$\Gamma^{\boldsymbol{m}}_{+\to-} = \pi g^2 \sum_{\boldsymbol{i}\in\text{edge}} C^2_{\boldsymbol{im}}|f_{\boldsymbol{i}}(E-\Delta_{\boldsymbol{m}}/2,-\Delta_{\boldsymbol{m}})|^2, \tag{4.1}$$

where $f_{\boldsymbol{i}}$ is the structure factor for the off-diagonal elements of the boundary operator $\gamma^x_{\boldsymbol{i}}$ and $\Delta_{\boldsymbol{m}}$ is the energy difference between the l-bit states $\tau'_{\boldsymbol{m}}$ and τ. The sum in Eq. 4.1 is dominated by the thermal spin at the boundary closest to $\boldsymbol{m}$ whose distance we denote by $R_{\boldsymbol{m}}$, and thus $\Gamma^{\boldsymbol{m}}_{+\to-} \sim \pi g^2 e^{-2R_{\boldsymbol{m}}/\xi}|f_{\boldsymbol{i}}(E-\Delta_{\boldsymbol{m}}/2,-\Delta_{\boldsymbol{m}})|^2$. The smallest rates are thus associated with the l-bits deep in the bulk, for which $R_{\boldsymbol{m}} \sim \alpha L$ and $\Gamma_{\min} \sim e^{-2\alpha L/\xi}$ where L is the linear dimension of the system. In order for the states $|E,\tau\rangle$, $|E',\tau'_{\boldsymbol{m}}\rangle$ to strongly mix, the rate of hybridization $\Gamma^{\boldsymbol{m}}_{+\to-}$ needs to exceed the many-body level spacing of the eigenstates $|E\rangle$ of the thermal boundary which scales as $\delta_E \sim 2^{-L^{d-1}}$. This statement is equivalent to checking hybridization from first order perturbation theory. As a result, for $d > 2$, the states strongly mix and for $d = 2$, which is a marginal case, a finite fraction of the l-bits mix, those within distance $R = \frac{\xi\ln(2)L}{2}$ of the edge. However, even for $d \geq 2$, where the l-bits thermalize in the infinite-time limit, the rate $\Gamma_{\min} \sim e^{-2\alpha L/\xi}$ is so small that the thermalization time is exponentially large, thus in the thermodynamic limit the full system is effectively MBL. Additionally, the l-bit picture and the localization properties of the eigenstates, which are thought of as defining features of the MBL phase, are no longer valid: the l-bits are only approximately conserved, in the sense the operator norm of the commutator $||[H,\tau^z_{\boldsymbol{i}}]|| \sim e^{-L/\xi}$ is exponentially small. The authors go further to show that

for $d > 2$ the existence of almost conserved l-bit operators τ_i^z is consistent with ETH [84], though the width of their structure factor is exponentially small in L.

In Ref. [85] the authors argue that any nonzero density of large enough thermalizing inclusions will destroy the MBL phase in $d > 1$, and in $d = 1$ if the localization length is larger than a finite threshold, converting it to an extremely slow, but ultimately thermalizing regime. The outline of their argument is as follows. Consider a single thermal inclusion. If the inclusion is sufficiently coupled to the MBL phase in which it is embedded, then it will thermalize the localized degrees of freedom (l-bits) bordering the inclusion, and this allows the thermal inclusion to grow in size. For concreteness, the inclusion is described by an ergodic Hamiltonian H_E that is coupled to a bordering l-bit with Hamiltonian $H_1 = h_1\tau_1^z$ by $H_{\text{int}}^{(1)} = g_1\sigma_0^x\tau_1^x$, where g_1 is the coupling strength and σ_0^x is a local operator on the thermal inclusion. The eigenstates of the uncoupled system are of the form $|E, s_1\rangle$, where $s_1 = \pm 1$. The parameter $\mathcal{G}$ determining if the inclusion hybridizes the spin, from first order perturbation theory, reads:

$$\mathcal{G} = \frac{g_1\langle E' = E + 2s_1h_1, -s_1|\sigma_0^x\tau_1^x|E, s_1\rangle}{\delta_E} \tag{4.2}$$

$$= \frac{g_1\langle E' = E + 2s_1h_1|\sigma_0^x|E\rangle}{\delta_E} \tag{4.3}$$

$$= g_1\sqrt{\rho_E}|f(E + s_1h_1, 2s_1h_1)|. \tag{4.4}$$

where $\delta_E = 1/\rho_E$ is the level spacing corresponding to the thermal inclusion and we have used the ETH to describe the off-diagonal elements of the local operator σ_0^x, i.e. $\langle E' = E + 2s_1h_1|\sigma_0^x|E\rangle = \frac{|f(E+s_1h_1,2s_1h_1)|}{\rho_E}$. The structure factor f with respect to the second argument has width w, the typical energy per spin in the system. If $\mathcal{G} \gg 1$, then the l-bit is hybridized. [85] further assumes the eigenstates of the combined system $|E_1\rangle$ are random in the basis $|E, s_1\rangle$, except for energy conservation, in the sense that their weight is restricted in the energy space of the uncoupled eigenstates to a window of size $w_1 = g_1^2/w$ around the mean energy (this can be justified based on the Fermi Golden Rule). Now, when a second neighbouring l-bit with $H_2 = h_2\tau_2^z$ is considered, which is similarly coupled by $H_{\text{int}}^{(2)} = g_2\sigma_0^x\tau_2^x$ to the inclusion, the hybridization parameter reads

$$\mathcal{G}_2 = \frac{g_2\langle E_1' = E_1 + 2s_2h_2, -s_2|\sigma_0^x\tau_2^x|E_1, s_2\rangle}{\delta_{E_1}} = \frac{g_2\langle E_1' = E_1 + 2s_2h_2|\sigma_0^x|E_1\rangle}{\delta_{E_1}} \tag{4.5}$$

The authors show that under the previous assumptions for the eigenstates $|E_1\rangle$ the off-diagonal elements in Eq. 4.5 satisfy ETH, i.e.

$$\langle E_1' = E_1 + 2s_2h_2|\sigma_0^x|E_1\rangle = \frac{|f'(E + s_2h_2, 2s_2h_2)|}{\rho_{E_1}}, \tag{4.6}$$

for the combined system formed by the thermal inclusion and the first l-bit, where $\rho_{E_1} = 2\rho_E$ is the density of states for the combined system. However, the structure factor f' broadens with respect to the second argument by $w_1 = g_1^2/w$ compared to the original one, f, but this change is small since $g_1 \ll w$ in the MBL phase. Hence, the hybridization condition for l-bit 2, $\mathcal{G}_2 = g_2\sqrt{\rho_{E_1}}|f'(E+s_2h_2, 2s_2h_2)|$ is the same as if the original inclusion had grown by one spin. This is the avalanche effect whereby thermalized l-bits effectively enhance the original thermal degrees of freedom.

The previous analysis was then applied to MBL models with random disorder in any dimensions to the scenario wherein a spherical thermal inclusion is enclosed by l-bits which interact with it with strength $g_{\boldsymbol{i}} \sim e^{-R_{\boldsymbol{i}}/\xi}$ depending on their distance to the inclusion $R_{\boldsymbol{i}}$. It was found that a single thermal inclusion can destroy the MBL phase in $d > 1$ and in $d = 1$ if ξ is larger than a finite threshold. The latter was recently demonstrated numerically in Ref. [88]. As a result, there is a finite, but possibly extremely large time-scale over which a putative MBL phase is thermalized by the nonzero density of rare thermal inclusions which arise for any strength of disorder [85]. This time scale can be estimated from the Fermi Golden Rule and scales as $\sim w/g_{\boldsymbol{i}}^2$. Since the $g_{\boldsymbol{i}}$ decay exponentially with the distance from the l-bits to the inclusion, this leads to *very* long thermalization times. This argument leaves open the possibility that MBL phases can remain stable in higher d in nonrandom systems with quasiperiodic fields or other potentials that do not produce such rare thermalizing regions.

4.3 Model

Our Hamiltonian for a thermal inclusion connected to N l-bits is:

$$H = \sum_i \tilde{\Delta}_i \tau_i^z + J \sum_i (A_i \tau_i^z + B_i \tau_i^x) \tag{4.7}$$

where $\tilde{\Delta}_i$ are independently sampled random variables drawn from the uniform distribution on the interval $[-W, W]$, τ_i^α for $\alpha = x, y, z$ are the Pauli operators of l-bit i, and A_i and B_i are real Gaussian symmetric random matrices (GOE) acting on the thermal inclusion Hilbert space of dimension 2^M. We caution the reader that while the τ_i operators act on l-bit i, the A_i and B_i operators *act* on the central spin; the subscript i indicates which l-bit they are coupled to in H. The normalization of the GOE matrices is such that the off-diagonal elements have variance $1/2^M$, while diagonal elements have twice the variance. This normalization guarantees that the operator norm of A_i, B_i is order 1, which is the appropriate scaling for a local operator acting on the inclusion.

Here we focus on the simplest case of *a central spin model* with $M = 1$, as it captures most of the physics of the general $M > 1$ model in Eq. 4.7. At $M = 1$, it is convenient to expand the A_i matrices in the Pauli basis σ of the central spin:

$$A_i = a_i^0 + a_i^x \sigma^x + a_i^y \sigma^y + a_i^z \sigma^z \tag{4.8}$$

As A_i is a GOE matrix, $a_i^y = 0$, while a_i^α for $\alpha = 0, x, z$ are independent Gaussian random variables with zero mean and variance $1/2$. The Hamiltonian for the central spin model is therefore:

$$H = \sum_{i=1}^{N} J(\mathbf{a}_i \cdot \sigma)\tau_i^z + JB_i\tau_i^x + \Delta_i \tau_i^z, \tag{4.9}$$

where $\Delta_i = \tilde{\Delta}_i + Ja_i^0$ is a renormalized field.

There are two dimensionless scales in the model: the temperature and the coupling constant J in units of W. As here we will focus on infinite temperature, J/W determines the entire dynamical phase diagram. At $J = 0$, none of the spins are coupled and the eigenstates are trivially localized product states with $\tau_i^z = \pm 1$. Each τ^z state of the l-bits is doubly degenerate as the energy is independent of the state of the central spin σ. In anticipation of the leading interaction term at non-zero J, we choose to work in a basis in which σ points along or against the effective τ_i^z-dependent field:

$$\mathbf{h}(\tau) = J \sum_i \mathbf{a}_i \tau_i^z \tag{4.10}$$

Thus, $|\, s\hat{h}, \tau\rangle$ for $s = \pm 1$ is an eigenstate of Eq. 4.9 with $J = 0$, and, in fact, even for $J \neq 0$ for $B_i = 0$, i.e. ignoring the interaction terms $JB_i\tau_i^x$ which flip the state of the l-bits τ_i^z. This set of 'Fock states' are naturally viewed as the vertices of a $(N+1)$-dimensional hypercube. We use the shorthand $\tau = (\tau_1^z, \ldots, \tau_i^z, \ldots, \tau_N^z)$ and drop the explicit dependence of $\mathbf{h}$ on τ wherever it is clear from context.

The statistics of the effective field $\mathbf{h}$ play an important role in the dynamics of the model. As the sum of N independent Gaussian random vectors of $O(J)$ in the xz plane, $\mathbf{h}$ has mean zero and typical (root-mean-square) length $J\sqrt{N}$ by the central limit theorem or from the addivity of the variance for the sum of independent Gaussian random variables. Its distribution is Gaussian with respect to disorder fluctuations at fixed τ^z (by construction as it is a sum of random Gaussian variables) and with respect to varying τ_i^z in a fixed sample (by the central limiting theorem).

In the opposite limit of $J \gg W$, the bare random fields $\tilde{\Delta}_i$ become negligible. Each l-bit feels a random field of order J in the xz plane and interacts with the central spin

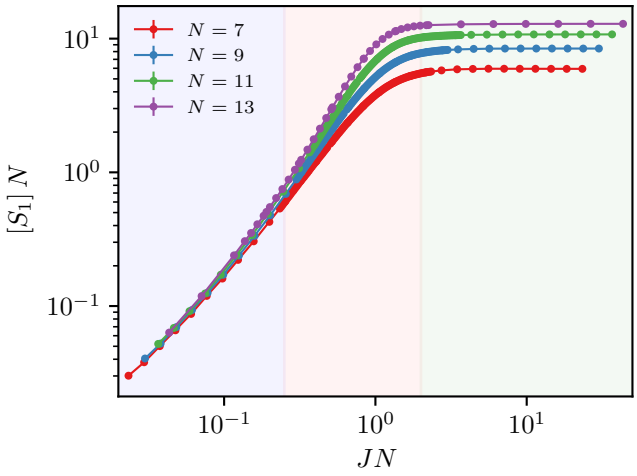

Figure 4.2: Scaled mean single l-bit entanglement entropy $[S_1]N$ vs the scaled interaction strength JN for various N. There are three qualitative regimes: small $JN \lesssim 0.2$, where the curves collapse, consistent with MBL; large $JN \gtrsim 1$, where N dependent plateaus form, consistent with the approach to thermalization; and a wide crossover regime at intermediate JN.

with strength order J. As the disorder and the interactions are of comparable strength, we expect the system to thermalize as $N \to \infty$. Both the perturbative (in B) arguments and the numerical exact diagonalization study discussed below confirm this expectation. However, as W drops out for large J, we cannot simply increase J in order to approach a more robust thermalizing limit at finite size. This unfortunately results in large finite-size effects in the thermalizing regime.

4.4 Numerical results

We study Eq. 4.9 using full diagonalization for a $M = 1$ central spin coupled to $N = 7$ to $N = 13$ l-bits. The number of samples at each (N, J) is 2500, except for $N = 13$ where this number is 600, and within each sample, we restrict our analysis to the eigenstates within the central half of the energy spectrum in order to study the properties of infinite temperature. The mean of a quantity q is denoted by $[q]$, while the standard deviation is denoted by Δq. Unless specified otherwise, the mean and standard deviation are taken with respect to all the l-bits (l), the eigenstates in the central half of the spectrum (E),

and all samples (s). When the statistical operation is restricted to a particular subset of these, we include the subset in the subscript. We measure energy in units where $W = 1$.

4.4.1 Three Regimes

Numerically, we find three distinct regimes as a function of rescaled coupling JN: localized (MBL) at small JN, ergodic (ETH) at sufficiently large JN and a wide crossover at intermediate JN. This is conveniently summarized by the behavior of the mean single l-bit entanglement entropy $[S_1]$ (see Fig. 4.2). The entropy, S_1^i, measures the degree to which l-bit i is thermalized within an eigenstate $|E_m\rangle$:

$$S_1^i = -\mathrm{Tr}\rho_i \log_2 \rho_i \tag{4.11}$$

Here, $\rho_i = \mathrm{Tr}_{\bar{i}}|E_m\rangle\langle E_m|$ is the reduced density matrix for site i, where $\bar{i}$ denotes the complement of i. If the state $|E_m\rangle$ is thermal (at infinite temperature) for l-bit i, then $S_1^i = 1$ obtains its maximal value. In a localized state, on the other hand, S_1^i can be < 1.

The entropy shows the three qualitatively distinct regimes in Fig. 4.2. At sufficiently small coupling, we find that $[S_1]N$ collapses onto a single curve to excellent precision. This strongly localized behavior ($[S_1] \sim 1/N \to 0$ as $N \to \infty$) also follows in the perturbative treatment of the localized phase in Section 4.5. This is a particularly strong form of localization, seen also in other long-range models [89], as $[S_1] \sim O(1)$ in 1d MBL phases [31, 90]. We consider this regime the finite size precursor to the MBL phase. At sufficiently large coupling $JN \gtrsim 1$, $[S_1]N$ develops plateaus which increase with N, consistent with thermalizing behavior in the thermodynamic limit ($[S_1] \to 1$ as $N \to \infty$). Finally, the wide crossover between these two behaviors exhibits growth of $[S_1]N$ with N, which suggests at least partial delocalization.

In order to focus in on the intermediate transition regime, we look at the finite size behavior of $[S_1]$ (unscaled) along with that of the mean level spacing ratio $[r]$ (see Fig. 4.3) The level spacing ratio r is defined as the ratio of consecutive level spacings $r_m = \frac{\min(s_m, s_{m+1})}{\max(s_m, s_{m+1})}$ with $s_m = E_m - E_{m-1}$ when the energies E_m are enumerated in increasing order. The ratio measures the level repulsion in a system and is commonly used to diagnose (de)-localization. It flows to $r_{\mathrm{GOE}} \approx 0.53$ in ergodic systems that are characterized by RMT statistics and to $r_{\mathrm{Poisson}} \approx 0.39$ for systems with Poisson level statistics (c.f. Section 1.3.1). The upper panel of Fig. 4.3 indicates that $[S_1]$ has a sharpening crossover from 0 to 1 near $JN \sim 0.5$ while the lower panel shows a similar sharpening crossover in $[r]$ at $JN \sim 1$. This suggests that the crossover between the localized and thermal phases sharpens into a

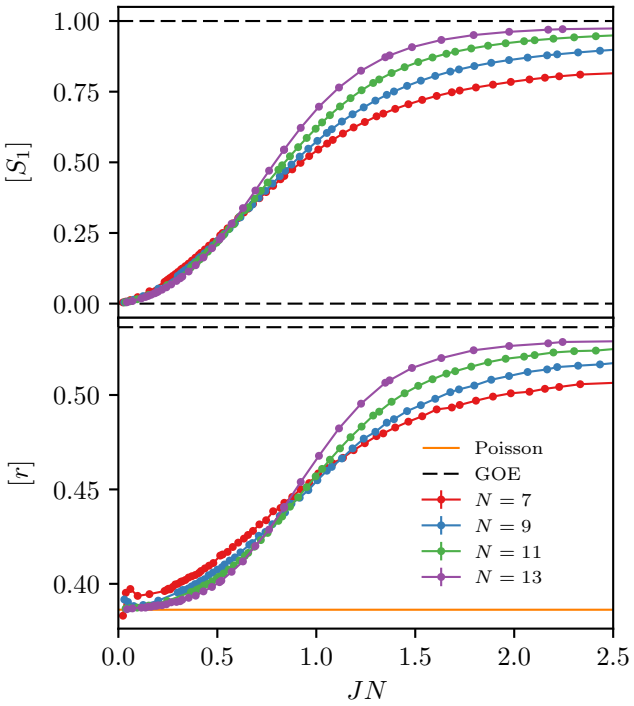

Figure 4.3: (top) Mean single l-bit entanglement entropy $[S_1]$ vs the scaled interaction strength JN and, (bottom) mean level spacing ratio $[r]$ vs JN, at different N. Both plots show sharpening of the transition as N is increased, but with significant finite size "drift" of the crossing points, particularly for $[r]$.

phase transition on the scale JN. However, as the finite size effects are large, it is difficult to separate two possible scenarios: The first scenario posits that the two crossover points coalesce as $N \to \infty$ and there is a direct transition from MBL to a fully delocalized thermal phase on the scale JN (up to smaller corrections, e.g. a multiplicative logarithmic [91]). In the second scenario, the crossover points remain separate as $N \to \infty$, sharpening into two phase transitions surrounded by an intervening partially delocalized phase [91]. In this phase, the eigenstates overlap with an exponentially large subset of the basis states, but are not fully delocalized in the Hilbert space. As a result, one would expect that $[S_1] = 1$, the level statistics are Poisson, and it would presumably not satisfy ETH (in particular it would violate the ETH prediction for the scaling of the off-diagonal elements and the fluctuations of the diagonal elements of local operators in the eigenstate basis). This is often referred to as a delocalized non-ergodic phase.

We comment that $[S_1]$ and $[r]$ both attain plateaus at large JN in the thermal phase.

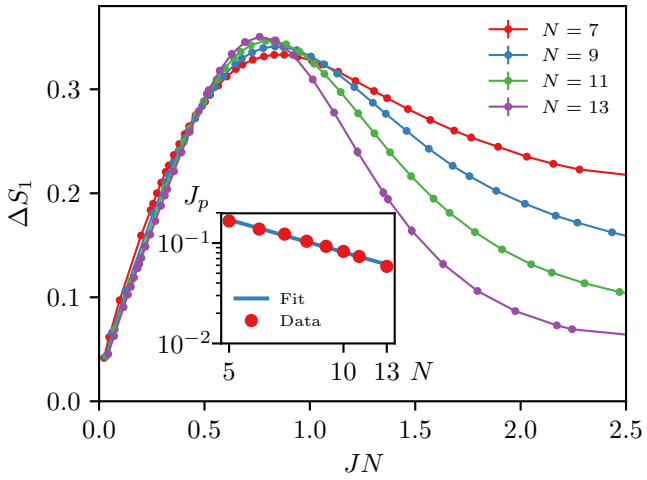

Figure 4.4: Standard deviation of S_1 over l-bits, states and samples ΔS_1 vs scaled coupling JN. (inset) Scaling of peak position J_p with N on log-log scale. Least squares fit (blue) to $J_p \approx 0.95 N^{-1.07}$.

We have checked that the plateau values approach their limiting ETH values as $2^{-N/2}$.

4.4.2 Heterogeneity in Observables

The most striking feature of the crossover regime is the extreme heterogeneity of single site observables across eigenstates, samples and l-bits. As a coarse measure, consider the standard deviation of the single site entanglement entropy ΔS_1 shown in Fig. 4.4. In both the MBL and ETH regimes of Fig. 4.4, the "flow" of ΔS_1 is to 0 as N grows. This is to be expected in the infinite temperature ETH phase, where the fluctuations in single-site observables across eigenstates are exponentially small in N ($\Delta S_1 \sim 2^{-N/2}$ to be precise). On the MBL side, the perturbative picture of Section 4.5 indicates that the entire distribution of S_1 scales to 0 as $1/N$ so that $\Delta S_1 \sim 1/N$ as well. In contrast, in the one dimensional MBL phase, ΔS_1 is non-zero as $N \to \infty$. This can be seen explicitly in, for example, the 1d phenomenological models of [25, 26], where the Hamiltonian in the MBL phase can be cast in a diagonal form in terms of l-bits, which are dressed versions of the original (physical) bits with finite localization length (c.f. Eq. 1.20 and related discussion). The data in Fig. 4.4 is consistent with the previous predictions from the perturbative picture.

In the crossover regime, however, ΔS_1 shows a peak whose height ΔS_1^p increases with N.

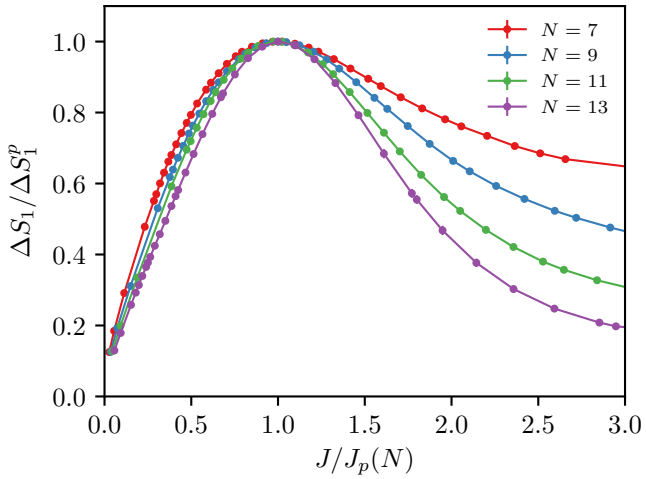

Figure 4.5: $\Delta S_1/\Delta S_1^p$ narrows on the scale $J/J_p(N)$ with increasing system size. This is consistent with the finite size crossover sharpening into a delocalization phase transition in the limit of large N.

As S_1 is a bounded variable, the growth of ΔS_1^p must saturate at larger sizes. Nonetheless, should the peak persist to the thermodynamic limit, it must converge to a critical point (see the discussion below). Previous works in one dimensional models have likewise used the peak as a sensitive proxy for the critical point [73, 87]. We accordingly define $J_p(N)$ by the location of the peak in ΔS_1 and study the properties of the critical region by following this coupling. The inset to Fig. 4.4 confirms that it scales as $J_p(N) \sim W/N$.

The relative width of the peak at $J = J_p(N)$ narrows as N increases (see Fig. 4.5). This trend is consistent with the existence of a sharp phase transition in S_1 at $J_p(N)$ in the thermodynamic limit. Indeed, $J = J_p(N)$ approaches the location of the crossing point in $[S_1]$ in Fig. 4.3(a) at $JN \approx 0.5$ with increasing N. As S_1 detects if l-bits are entangled in eigenstates, we expect that it is sensitive to whether the system is localized or delocalized on the classical hypercube of 'Fock states', but not necessarily to whether the delocalized phase satisfies the off-diagonal criteria for ETH. Fig. 4.5 is therefore consistent with both scenarios for the phase diagram discussed in Section 4.4.1. In the first scenario, there is a direct phase transition from a MBL to an ETH phase at $J = J_p(N)$ as $N \to \infty$. In the second scenario in where there are two phase transitions with an intervening partially delocalized phase, $J = J_p(N)$ is the position of the first transition out of the localized

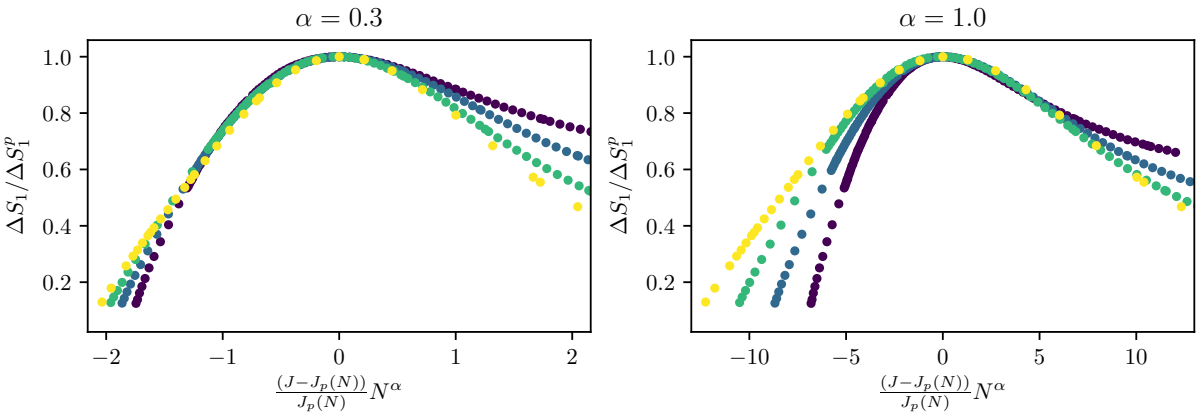

Figure 4.6: Finite size scaling of the ratio $\Delta S_1/\Delta S_1^p = f(\frac{J-J_p(N)}{J_p}N^\alpha)$ with $\alpha = 0.3$ and $\alpha = 1.0$, showing a good collapse of the data in the localized and delocalized phase, respectively.

phase.

We note that the narrowing of this peak on the localized side is slower than on the delocalized side. This can be clearly seen in Fig. 4.6, where the $\Delta S_1/\Delta S_1^p$ data on the localized narrows on the scale of $J_p(N)$ as N^α with $\alpha \sim 0.3$, while it narrows faster in the delocalized side with $\alpha \sim 1.0$. This makes the extraction of a finite size scaling exponent problematic, and indicates that the numerics are not yet in the asymptotic scaling regime as we discuss below. Moreover, the narrowing on the delocalized side suggests exponents which violate finite size scaling bounds [92, 93].

We now characterize the heterogeneity in S_1 in the crossover region in more detail. Fig. 4.7(a) shows the wide variation of S_1 across both l-bits and eigenstates within a typical sample. More quantitatively, Fig. 4.7(b) shows that the distribution of S_1 across l-bits, eigenstates and samples at $J = J_p(N)$ is increasingly bimodal with increasing N, with increasing weight near zero and one and decreasing weight at intermediate values of S_1. Thus, l-bits are mostly either 'on' (highly entangled) or 'off' in any given eigenstate in the crossover regime. The bimodal distribution underlies the large peak in ΔS_1 seen in Fig. 4.4.

As l-bits in weak (strong) fields might strongly (weakly) entangle with the central spin and with one another, the reader may not be surprised by the bimodality seen at J_p. This explanation however misses the remarkable feature that *different* l-bits are active in different eigenstates. To demonstrate this effect, we investigate the sample averaged variation of S_1^i across eigenstates, $[\Delta_E S_1^i]_s$, for fixed l-bit i (see Fig. 4.8(a)). If l-bit i were

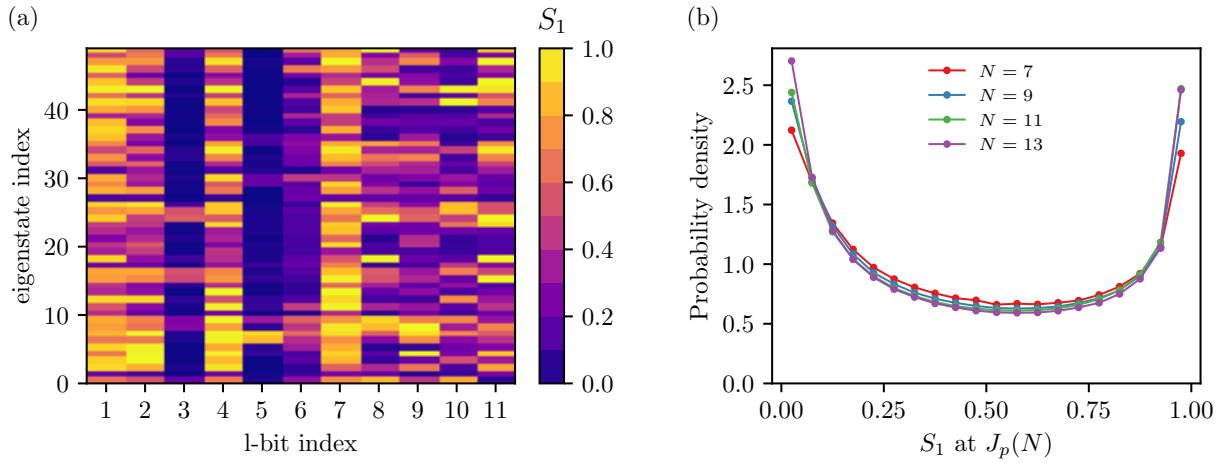

Figure 4.7: (a) S_1 heatmap from a specific sample of the central spin model with $N = 11$ l-bits at the peak coupling J_p. Each row corresponds to an eigenstate from 50 infinite temperature selected eigenstates and each column to a particular l-bit. (b) Histogram of S_1 collected across l-bits, states and samples at J_p.

entangled/unentangled across all the states, $[\Delta_E S_1^i]_s$ would be small. Instead, we see a robust peak in this quantity at $J = J_p(N)$ at each N, with a magnitude comparable to the total variation of S_1 across all samples, states and l-bits (ΔS_1).

In the remainder of Fig. 4.8, we parse the contributions to ΔS_1 coming from sample, eigenstate and l-bit fluctuations, all of which contribute significantly to the total variation at these sizes [87]. We find that for our range of N the largest contributions to ΔS_1 near the transition come from variations between l-bits in the same eigenstate, i.e. $[\Delta_l S_1]_{E,s}$ (panel (b)), and across different eigenstates for the same l-bit, i.e. $[\Delta_E S_1^i]_s$ (panel (a)). This is in contrast with previously studied 1d models with quenched disorder exhibiting an MBL-thermal transition: in Ref. [87], the sample-to-sample variation is the dominant contribution at the largest sizes accessible to diagonalization.

The sample-to-sample variation shown in panel (d) is the only component which shows an *accelerating* trend to stronger and sharper peaks as N increases. This increasing trend cannot continue with increasing N as ΔS_1 is bounded; this indicates that we are not in the asymptotic scaling regime at the numerically accessible system sizes. Nevertheless, should the sample-to-sample variation become the dominant contribution to ΔS in the thermodynamic limit, the transition is first order in the sense that we discuss in Section 4.6. The inhomogeneity between eigenstates in the l-bit-averaged entanglement at the transition is shown in panel (c). The peak shows a slight decrease with N. If this persists to larger N, then all of the eigenstates in one sample would become similar in this respect although

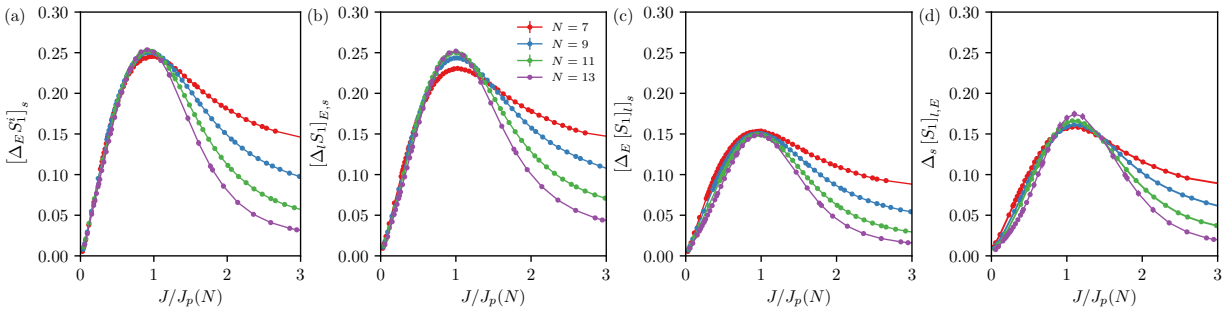

Figure 4.8: Parsing of contributions to the fluctuations of the single l-bit entanglement entropy S_1 vs scaled J. (a) Standard deviation of S_1^i for a fixed l-bit i over eigenstates, averaged over samples. (b) Standard deviation of S_1 over l-bits, averaged over eigenstates and samples. (c) Average over samples of the standard deviation over eigenstates of the l-bit-averaged S_1. (d) Standard deviation over samples of S_1 averaged over eigenstates and l-bits.

it seems likely that they will still differ in which l-bits are less versus more entangled.

Finally, we note that this central spin model exhibits the strongest single-site bimodality in the MBL-ETH crossover of any numerically studied model to date, suggesting that the transition may indeed be 'first order' [86]. As we would expect to see the probability distribution of S_1 approach a delta-function at $S_1 = 1$ (the required value in the thermal phase) if the transition was continuous. Furthermore, Fig. 4.8(a) shows that S_1^i fluctuates strongly between eigenstates of a single sample, even at infinite temperature. No renormalization group treatment to date takes these intra-sample variations between eigenstates into account, so might be missing some important physics of this transition [94, 95].

4.4.3 Eigenstate Distributions

As discussed in Section 4.3, the eigenstates at $JN = 0$ are the product states $|s\hat{h}, \tau\rangle$. It is therefore natural to study the support of the eigenstates at $JN > 0$ on this basis (which forms a hypercube). This is measured by the participation ratio PR of an eigenstate $|E_m\rangle$ (c.f. Section 1.3.3):

$$\mathrm{PR}^{-1}(E_m) = \sum_{s,\tau} |\langle E_m | s\hat{h}, \tau\rangle|^4 \tag{4.12}$$

Fig. 4.9 shows the evolution of the distribution of the participation ratios with coupling.

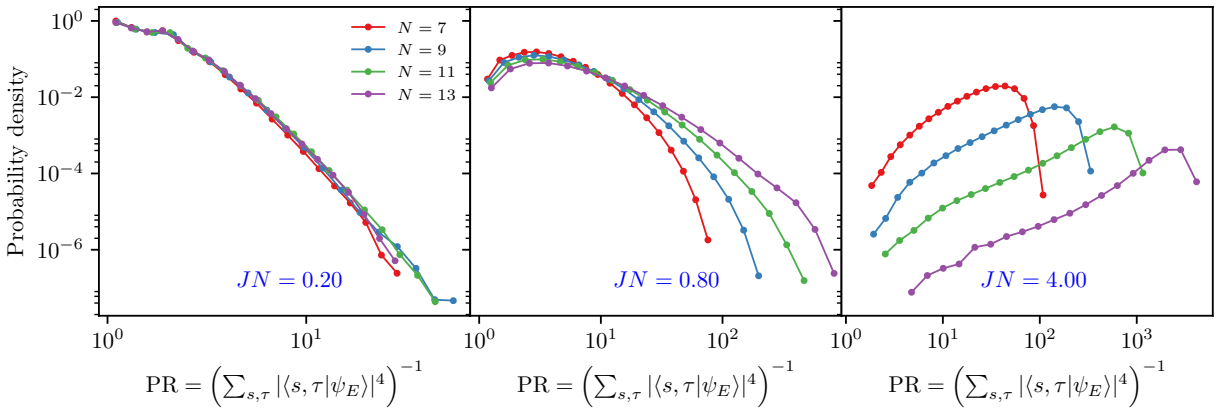

Figure 4.9: Probability density function of the eigenstate participation ratio (PR) on the classical basis $|s\hat{h}, \tau\rangle$ for JN in the three different regimes: deep in the MBL phase ($JN = 0.2$), in the crossover region ($JN = 0.8$), and in the thermal phase ($JN = 4.0$).

In the localized phase (illustrated by $JN = 0.2$), the distribution strongly decays with PR and is independent of N. This implies that the eigenstates are essentially localized to finite regions of the hypercube even in the thermodynamic limit. This is consistent with the perturbative picture of the localized phase in Section 4.5.

In the crossover regime (at $JN = 0.8$), the distribution remains peaked at small PR, but the weight in the tail grows with N. Further, the scale of the cut-off grows exponentially with N. In the second order perturbative treatment developed in Section 4.5, the transition coincides with the high-dimensional percolation transition on the hypercube. In this picture, the PR distribution should decay as a power law with exponent 3/2 which is cutoff by $n_\xi \sim 2^{2N/3}$. This (exponential) scaling of the cutoff is in qualitative agreement with the numerical data at $JN = 0.8$, although the curvature in the bulk of the distribution is clearly inconsistent with a simple power law.

Finally, deep in the ergodic phase (at $JN = 4$), the PR at the peak of the distribution increases exponentially with N, consistent with the ETH. The probability density at finite PR appears to decrease exponentially with N. This would be consistent with exponentially rare samples probing the localized side of the transition.

4.4.4 Central spin entanglement

We now study the behavior of the central spin, which has not been addressed up to this point. In particular, we have analyzed the behaviour of its entanglement S_{cs}, in eigenstates,

with the rest of system, i.e. with all the l-bits. At $B_i = 0$, the eigenstates $|s\hat{h}(\tau), \tau\rangle$ are product states between the central spin and the l-bits with $S_{\rm cs} = 0$ in all eigenstates and samples. At small JB_i, the central spin can flip in the resonant processes that hybridize the classical states in first order perturbation theory (see the perturbative analysis in Section 4.5). As the central spin is substantially entangled even in eigenstates that are spread over a just few classical states, we expect that $S_{\rm cs}$ is close to one in eigenstates with even a few resonant l-bits, while it is close to zero in eigenstates with no resonant l-bits. This has two important consequences: (1) the moments of the distribution of $S_{\rm cs}$ are non-zero in the MBL phase even in the thermodynamic limit, and (2) the central spin crosses over to high entanglement at $J_p^{\rm cs}(N) < J_p(N)$ within the MBL phase. Note however that the scaling of $J_p^{\rm cs}(N)$ with N is the same as that of $J_p(N)$ that is, $J_p^{\rm cs}(N) \sim W/N$. This can be seen in Fig. 4.11(b), where the fit $J_p^{\rm cs} \approx 0.17 N^{-1.18}$ is in qualitative agreement with this hypothesis especially considering the limited systems sizes analysed.

Feature (1) is already visible in the mean $[S_{\rm cs}]$ of the central spin entanglement (see Fig. 4.10(a)) where the N dependence only emerges for $J \sim J_p(N)$. The total standard deviation $\Delta S_{\rm cs}$ brings out feature (2) more clearly (see Fig. 4.10(b)) as the peak lies at $J \sim J_p^{\rm cs}(N) < J_p(N)$ to the left of the global transition point. Furthermore, the peak value of $\Delta S_{\rm cs}$ exceeds 0.28 (the standard deviation of a uniform distribution in the interval $[0, 1]$), showing that the distribution is bimodal in the vicinity of $J_p^{\rm cs}(N)$ and this is confirmed in Fig. 4.11(a) which shows the probability density of this quantity. Finally, in the ergodic phase at $J/J_p(N) > 1$, $\Delta S_{\rm cs}$ decreases with increasing N, in agreement with the ETH.

In order to bring out the sharp contrast between ΔS_1 of the l-bits (Fig. 4.5) and $\Delta S_{\rm cs}$ of the central spin, we plot $\Delta S_{\rm cs}$ normalized by its peak value vs $J/J_p^{\rm cs}(N)$ in Fig. 4.10(b). Unlike the l-bit entanglement entropy variation, $\Delta S_{\rm cs}$ does not narrow on its own scale with increasing N in the MBL phase. Hence the central spin's entanglement undergoes a crossover within the MBL phase, as opposed to the l-bits for which the peaks narrow as N is increased as seen in Fig. 4.5.

4.5 Perturbative analysis

The central spin model is amenable to perturbative study in small B relative to the "classical" configurations $|\, s\hat{h}, \tau\rangle$ defined in Section 4.3. In this approach, delocalization takes place when a typical starting configuration resonates with a divergent subgraph of degenerate configurations on the hypercube.

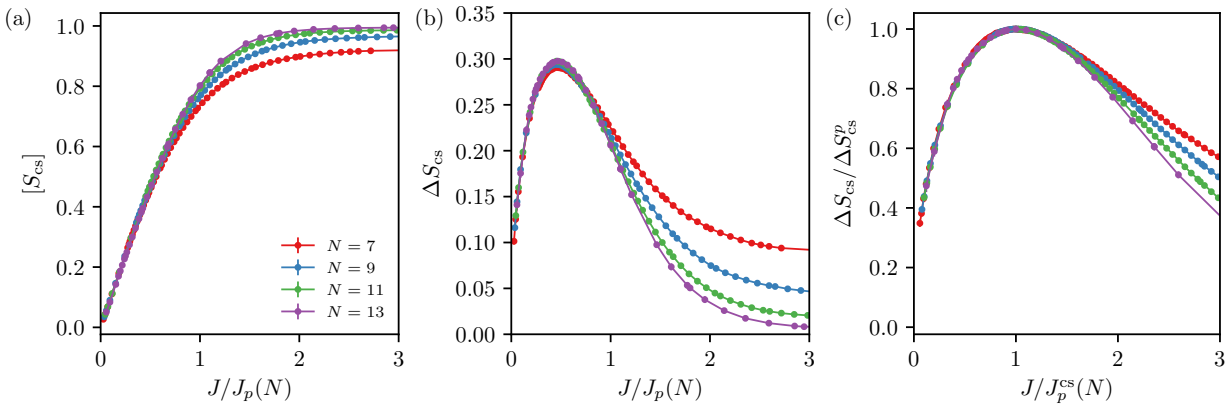

Figure 4.10: (a) Mean entanglement entropy of the central spin $[S_{\rm cs}]$ and (b) standard deviation $\Delta S_{\rm cs}$ vs. scaled coupling $J/J_p(N)$. (c) $\Delta S_{\rm cs}$ normalized by its value at the peak for a given number of l-bits N vs the coupling J on the scale of the coupling $J_p^{\rm cs}(N)$.

In Sections 4.5.2 and 4.5.3, we consider first and second order processes and show that a delocalization transition takes place at $J_c \sim W/N$. At this coupling, only a small (finite) number of spins can resonate by first order spin-flip processes. On the other hand, second order 'flip-flop' processes, in which two l-bits and the central spin are flipped, produce a divergent subgraph of resonant configurations for $J > J_c$. Thus, the low order perturbative analysis captures a delocalization transition on a scale $J_c \sim W/N$, consistent with the numerical observations.

As the resonant subgraphs produced at second order are locally tree-like, i.e. typically different pairs of l-bits are found to be resonant at each step, with an expected branching number $K \sim J^2N^2/W^2$, their statistical properties can be determined from independent second neighbor bond percolation on the hypercube with probability of placing a bond, $p \sim J^2/W^2$. We summarize the relevant properties of this percolation problem in Section 4.5.4. In Section 4.5.5, assuming that actual many-body eigenstates are simply delocalized over these resonant subgraphs, we explain the numerical observations that $[S_1] \sim 1/N$ in the localized phase, as well as the distribution of the participation ratios in the classical basis. Thus, the low order perturbative analysis also quantitatively describes the localized phase for $J < J_c$.

However, the low order perturbative analysis does not capture all of the features of the crossover region at $J \approx J_c$. This is not surprising as we expect higher order processes to be important in the vicinity of the delocalization transition. In Section 4.5.6, we adapt the arguments of Ref. [91] that suggest a transition from localized to a delocalized non-ergodic phase on the scale $J \sim W/N \ln N$. As discussed in Section 4.4, the numerical evidence for

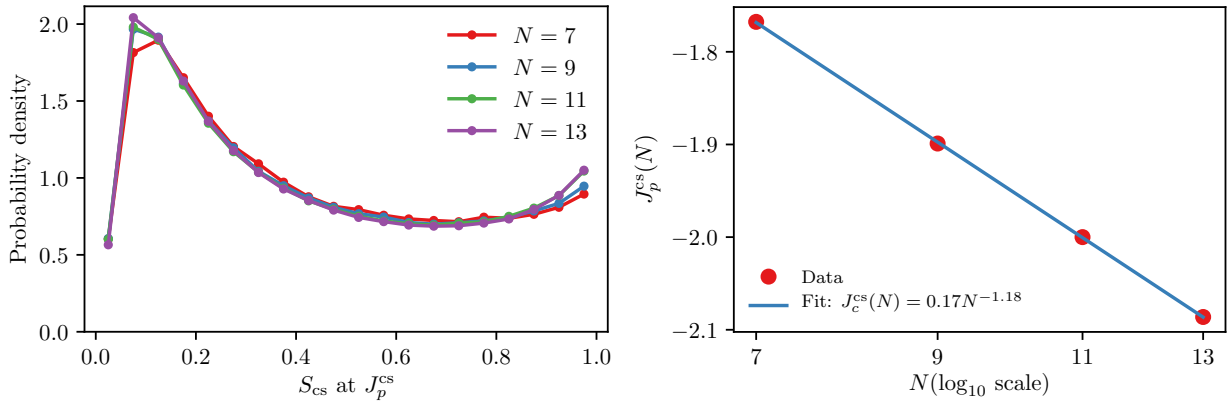

Figure 4.11: (a) Histogram of the entanglement entropy of the central spin [$S_{\rm cs}$] collected over different eigenstates and samples. (b) Scaling of peak position $J_p^{\rm cs}(N)$ with N on log-log scale. Least squares fit (blue) to $J_p^{\rm cs} \approx 0.17N^{-1.18}$.

this intervening phase is inconclusive.

4.5.1 Preliminaries

At $B = 0$, the state $|s\hat{h}, \tau\rangle$ has classical energy

$$E_{s\tau} = s|\mathbf{h}(\tau)| + \sum_{i=1}^{N} \Delta_i \tau_i^z \tag{4.13}$$

These $2^{(N+1)}$ configurations are naturally viewed as the corners of a hypercube. The interaction term, $J\sum_i B_i \tau_i^x$, defines a short range hopping model on the hypercube as it connects the corners related by single l-bit flips either with or without flipping the central spin s. As the perturbative arguments will predict a delocalization transition at $J_c \sim W/N \ll W$, we assume $J \ll W$ henceforth, since here we are only considering the large-N regime.

To perturb around a typical infinite temperature configuration $|s\hat{h}, \tau\rangle$, we need the classical energies of nearby configurations. We denote the l-bit configuration obtained by flipping k l-bits $i_1, \ldots, i_k$ by $\tau'_{i_1 \ldots i_k} = (\tau_1^z, \ldots, -\tau_{i_1}^z, \ldots, -\tau_{i_2}^z, \ldots, \tau_N^z)$. From Eq. 4.13, the energy of such a state relative to the initial state is

$$\Delta E_{s'\tau'_{i_1 \ldots i_k}} = s'|\mathbf{h}(\tau')| - s|\mathbf{h}(\tau)| - 2\sum_{m=1}^{k} \tau_{i_m}^z \Delta_{i_m} \tag{4.14}$$

For small numbers of flipped l-bits ($k \ll N$), the deviation in the effective field on the central spin,

$$J\delta_{i_1,\ldots i_M} \equiv \mathbf{h}(\tau') - \mathbf{h}(\tau) = -2\sum_{m=1}^{k} \mathbf{a}_{i_m}\tau^z_{i_m} \tag{4.15}$$

is much smaller than the field itself, $|\delta_{i_1,\ldots i_M}| \sim \sqrt{k} \ll \sqrt{N}$. Thus, expanding Eq. 4.14 in $J\delta/|\mathbf{h}(\tau)|$ (and suppressing the indices of the k flipped l-bits),

$$\begin{aligned}\Delta E_{s'\tau'} = (s'-s)|\mathbf{h}| - 2\sum_{m=1}^{k}\tau^z_{i_m}(\Delta_{i_m} + s'J\mathbf{a}_{i_m}\cdot\hat{h}) \\ + s'\frac{J^2}{2|\mathbf{h}|}\left(|\delta|^2 - (\hat{h}\cdot\delta)^2\right) + O\left(\frac{J^3}{|\mathbf{h}|^2}\right)\end{aligned} \tag{4.16}$$

The first term is the dominant energy change on flipping the central spin relative to its local field. It is of order $J\sqrt{N}$ (because it results from the sum of N random terms) but independent of the choice of flipped l-bits. The second term is the total field energy of the flipped l-bits. The bare fields $\Delta_i \sim W$ while the correction due to the central spin is only $\sim J \ll W$ (for J near the transition). Finally, the third and higher order terms represent the interactions between the flipped l-bits induced by their interaction with the central spin. In particular, the leading interaction $\sim \frac{J^2}{2|\mathbf{h}|}$ is of order $J/\sqrt{N}$, since $|\mathbf{h}| \sim J\sqrt{N}$.

4.5.2 First order processes

At first order in B, there are $2N$ neighboring configurations with a single l-bit flipped considering both flipping and not flipping the central spin. The corresponding matrix element is

$$J\langle s'\hat{h}(\tau'_j)|B_j|s\hat{h}(\tau)\rangle \sim O(J), \tag{4.17}$$

since B_j is a random matrix acting on the central spin with $O(1)$ matrix elements. The energy differences $\Delta E_{s'\tau'_j}$ over different choices of the l-bit j to be flipped are distributed on a band of width W (because the largest term in Eq. 4.16 comes from the random field $\Delta_j \sim O(W)$). Consequently, over $2N$ possibilities, the smallest $\Delta E_{s'\tau'_j}$ is of order $\sim W/2N$, and similarly the level spacing between the sorted in ascending order $\Delta E_{s'\tau'_j}$ is of the same order. Since a resonance occurs whenever the energy mismatch $\Delta E_{s'\tau'_j}$ is less than the matrix element which is of order $O(J)$, it follows that there are typically

$\sim J/(W/2N)$ resonant neighbors from a starting configuration. Thus, first order processes begin to find $O(1)$ resonances when $J \sim W/N$. We call these 'step 1' first order resonances. These resonances are treated appropriately by degenerate perturbation theory, whereby the perturbation is diagonalized exactly on these resonant configurations generically producing new basis states delocalized over the resonant cluster. These in turn connect at first order to configurations on the hypercube with two l-bit flips compared to the starting configuration $|s\hat{h}, \tau\rangle$. For $J \sim W/N$, there are again $O(1)$ next neighbor states which satisfy the resonance condition ('step 2' resonances). However, as we will describe in more detail in the next paragraph, the interaction energy in Eq. 4.16 $\sim J/\sqrt{N} \sim W/N^{3/2}$ is much smaller than the level spacing W/N, and roughly the same l-bit flips that were resonant at step 1 are resonant at step 2.

More precisely, if l-bit flip i_1 is resonant at step 1, $\Delta E_{s'\tau'_{i_1}} < O(J)$, then l-bit i_2 can resonantly flip at step 2 only if $\Delta E_{s''\tau'_{i_1 i_2}} < O(J)$. This requires that the effective field on l-bit i_2, $2\left(\Delta_{i_2} + s''J\mathbf{a_{i_2}} \cdot \hat{h}\right)$ itself should be close to 0 (if $s'' = s$) or $|\mathbf{h}|$ (if $s'' = -s$) to an accuracy of $O(J)$. The resonance condition at step 2 has not been changed compared to step 1. For example, if at step 1, $\Delta E_{s\tau'_i}$ satisfied the resonance condition, then, at step 2, $\Delta E_{s\tau'_{ij}}$ only differs from $\Delta E_{s\tau'_j}$ due to the interaction terms between i and j which are of order $J/\sqrt{N} \approx W/N^{3/2}$ and are negligible on the energy scale $J \sim W/N$ defining a resonance. Likewise, at step 2, $\Delta E_{-s\tau'_{ij}}$ only differs from $\Delta E_{-s\tau'_j}$ by terms of order $O(J)$ which is the same as the slack in the resonance condition. Thus, if $\Delta E_{-s\tau'_{ij}}$ satisfies the resonance condition at step 2, $\Delta E_{-s\tau'_j}$ would have already satisfied it at step 1. So, in both these examples further steps do not find new resonances and this applies generically. As there are only $O(1)$ l-bits whose field energy are close to 0 or $|\mathbf{h}|$ to accuracy $O(J)$, even multi-step first order processes only produce a finite cluster (i.e. $O(N^0)$) of resonant configurations. See Fig. 4.5.2.

We note that in the previous analysis, we have neglected the $O(J)$ shifts in the reference energy $E_{s\tau}$ which arise due to the diagonalization over the resonant configurations following the prescription of degenerate perturbation theory. But, this is justified. Since these corrections are of the same order as the slack in the resonance condition $O(J)$, they do not modify the statistics of the resonant clusters generated.

The structure of the resonant clusters suggests that there will be eigenstate dependence of which l-bits resonantly flip at $J \sim W/N$. There is an $O(1)$ subset of the l-bits with fields $\Delta_i \sim O(J)$ which resonate across all eigenstates without flipping the central spin. Meanwhile, those $\sqrt{N}$ l-bits with larger fields $\Delta_i \sim J\sqrt{N}$ resonate in eigenstates for which the central spin field $|\mathbf{h}|$ matches Δ_i (up to $O(J)$). However, when this happens for an

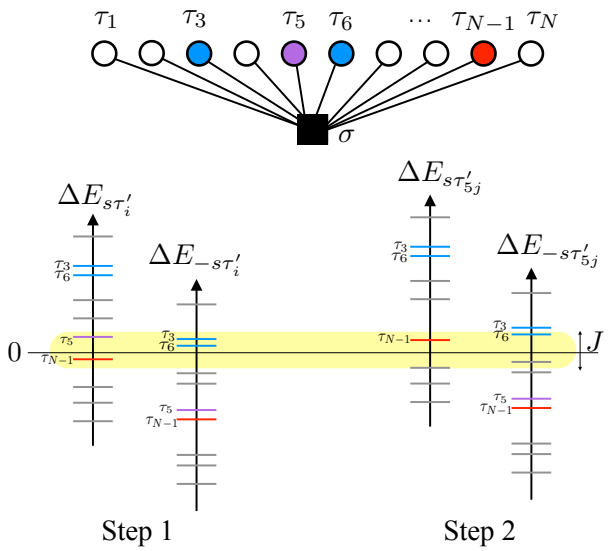

Figure 4.12: (a) Illustration of N l-bits coupled to the central spin σ. (b) Schematic of first order processes contributing to resonance relative to reference configuration $s\tau$. At step 1, there are N single l-bit flipped states with central spin s and N with central spin $-s$. Flipping the fifth (purple)/ (N-1)th (red) l-bits leads to purple/red states (also marked τ_5/τ_{N-1}) which satisfy the resonance condition $|\Delta E_{s\tau'}| < J$ without flipping the central spin; similarly, for the third and sixth l-bits (both in blue) with the central spin flipped ($-s$). Step 2 illustrates the levels accessible from the purple state (τ_5 flipped) at step 1. As the energy shifts are very small, roughly the same l-bit flips are resonant.

eigenstate, typically only $O(1)$ of these l-bits are resonant.

4.5.3 Second order processes

As all but $O(1)$ of the $2N$ nearest neighbor states are off-resonant for any particular state $|s\hat{h}(\tau)\rangle$ when $J \sim W/N$, we neglect in the following the first order resonances entirely.

At second order in perturbation theory in B, there are $2\binom{N}{2} \sim N^2$ configurations $|\, s'\hat{h}(\tau'_{ij}), \tau'_{ij}\rangle$, connected to an initial state $|\, s\hat{h}, \tau\rangle$. These correspond to flipping a pair of l-bits $i < j$ and, in addition possibly flipping the central spin (s') relative to its local field. For each final state, there are 4 channels contributing to the second order amplitude, corresponding to which l-bit (i or j) is flipped in the intermediate state and to whether

the central spin also flips in the intermediate state or not:

$$\langle s'\hat{h}(\tau'_{ij}),\tau'_{ij}\,|\,s\hat{h}(\tau),\tau\rangle^{(2)} = J^2\sum_{\tilde{s}=\pm1}\frac{\langle s'\hat{h}(\tau'_{ij})|B_i|\tilde{s}\hat{h}(\tau'_j)\rangle\langle\tilde{s}\hat{h}(\tau'_j)|B_j|s\hat{h}(\tau)\rangle}{(-\Delta E_{s'\tau'_{ij}})(-\Delta E_{\tilde{s}\tau'_j})} + (i\leftrightarrow j), \quad (4.18)$$

where $\Delta E_{s'\tau'_j} = E_{s'\tau'_j} - E_{s\tau}$ are the changes in energy relative to the initial state.

For each of these amplitudes, the final energy denominator, $\Delta E_{s'\tau'_{ij}}$, takes $\sim 2N^2$ values on a band of width W and thus has density of states $\sim 2N^2/W$. The intermediate energy differences, $\Delta E_{\tilde{s},\tau'}$, are typically $O(W)$ (we don't constrain in any way the energy of the intermediate state) and the numerators are $O(J^2)$, so large amplitudes (resonances) arise in Eq. 4.18 for $\sim (2N^2/W^2)J^2$ final states. This predicts $O(1)$ second order resonances when

$$J \sim W/N. \quad (4.19)$$

It turns out that this argument is essentially correct for those final states in which the central spin is flipped $s' = -s$. However, correlations among the matrix elements and energies involved in the four channels can result in destructive interference between the four amplitudes. As a result, this can lead to an enhancement of the required J to find of $O(1)$ second order resonances. Indeed, this is the case when the central spin is not flipped $s' = s$, as we show below.

To study the interference between the four second order channels it is simpler to consider pairs of channels first. For each pair, we label the intermediate states by 1 and 2, respectively, and the shared final state 12. The sum of amplitudes for the pair then reads

$$\frac{\mathcal{M}(1+\eta)}{(-\Delta E_{12})(-\Delta E_1)} + \frac{\mathcal{M}(1-\eta)}{(-\Delta E_{12})(-\Delta E_2)} \quad (4.20)$$

where $\Delta E_{12} = \Delta E_1 + \Delta E_2 + \Delta^{(12)}$ and $\mathcal{M}(1\pm\eta)$ are the numerators of the respective processes. The amplitude can be re-written as:

$$\frac{\mathcal{M}}{\Delta E_1 \Delta E_2}\left(1 - \frac{\Delta^{(12)}}{\Delta E_{12}} + \eta\frac{\Delta E_2 - \Delta E_1}{\Delta E_{12}}\right) \quad (4.21)$$

The first term $\frac{\mathcal{M}}{\Delta E_1 \Delta E_2}$ does not involve the final energy denominator ΔE_{12} at all, and thus, does not cause second order resonances (assuming that the first order differences $\Delta E_1, \Delta E_2$ are themselves not resonant which occurs for only $O(1)$ l-bits anyway). Resonances only arise due to interaction energy $\Delta^{(12)}$ and matrix element deviations

$\eta(\Delta E_2 - \Delta E_1)$. If both of these are parametrically small in N when defined in terms of disorder W, then resonances are suppressed at $J \sim W/N$, as to find a resonance we require $J^2N^2\Delta^{(12)}/W^3 \sim 1$ or $J^2N^2\eta(\Delta E_2 - \Delta E_1)/W^3 \sim 1$

Let us apply Eq. 4.21 to the determination of resonances in Eq. 4.18. There are two types of final states depending on whether the central spin is flipped ($s' = -s$) or not ($s' = s$) in the final configuration with respect to the initial one.

Central spin unflipped in the final state ($s' = s$)— In this case, we choose to analyze how the four channels destructively interfere in pairs according to the intermediate central spin state $\tilde{s}$. For completeness, we list here the four different amplitudes for flipping l-bits i, j.

$$J^2 \frac{\langle s\hat{h}(\tau'_{ij})|B_i|s\hat{h}(\tau'_j)\rangle\langle s\hat{h}(\tau'_j)|B_j|s\hat{h}(\tau)\rangle}{(-\Delta E_{s\tau'_{ij}})(-\Delta E_{s\tau'_j})} \tag{4.22}$$

$$J^2 \frac{\langle s\hat{h}(\tau'_{ij})|B_j|s\hat{h}(\tau'_i)\rangle\langle s\hat{h}(\tau'_i)|B_i|s\hat{h}(\tau)\rangle}{(-\Delta E_{s\tau'_{ij}})(-\Delta E_{s\tau'_i})} \tag{4.23}$$

$$J^2 \frac{\langle s\hat{h}(\tau'_{ij})|B_i| - s\hat{h}(\tau'_j)\rangle\langle -s\hat{h}(\tau'_j)|B_j|s\hat{h}(\tau)\rangle}{(-\Delta E_{s\tau'_{ij}})(-\Delta E_{-s\tau'_j})} \tag{4.24}$$

$$J^2 \frac{\langle s\hat{h}(\tau'_{ij})|B_j| - s\hat{h}(\tau'_i)\rangle\langle -s\hat{h}(\tau'_i)|B_i|s\hat{h}(\tau)\rangle}{(-\Delta E_{s\tau'_{ij}})(-\Delta E_{-s\tau'_i})} \tag{4.25}$$

For the pair of channels where the intermediate state has $\tilde{s} = s$, i.e. Eqs. 4.22, 4.23,

$$\Delta E_{s\tau'_{ij}} = \Delta E_{s\tau'_i} + \Delta E_{s\tau'_j} + s\frac{J^2}{|\mathbf{h}|}(\delta_i \cdot \delta_j - (\hat{h} \cdot \delta_i)(\hat{h} \cdot \delta_j)) + \cdots \tag{4.26}$$

so that

$$\Delta^{(ij)} = s\frac{J^2}{|\mathbf{h}|}(\delta_i \cdot \delta_j - (\hat{h} \cdot \delta_i)(\hat{h} \cdot \delta_j)) + \cdots = O\left(\frac{J}{\sqrt{N}}\right) \tag{4.27}$$

To leading order in $1/N$, the matrix elements are

$$\mathcal{M} = \frac{J^2}{2}\left(\langle s\hat{h}(\tau'_{ij})|B_i|s\hat{h}(\tau'_j)\rangle\langle s\hat{h}(\tau'_j)|B_j|s\hat{h}(\tau)\rangle + \langle s\hat{h}(\tau'_{ij})|B_j|s\hat{h}(\tau'_i)\rangle\langle s\hat{h}(\tau'_i)|B_i|s\hat{h}(\tau)\rangle\right) \tag{4.28}$$

$$= J^2 B_i^{++}B_j^{++} + O(J^2/\sqrt{N}) \tag{4.29}$$

$$\mathcal{M}\eta = \frac{J^2}{2}\left(\langle s\hat{h}(\tau'_{ij})|B_i|s\hat{h}(\tau'_j)\rangle\langle s\hat{h}(\tau'_j)|B_j|s\hat{h}(\tau)\rangle - \langle s\hat{h}(\tau'_{ij})|B_j|s\hat{h}(\tau'_i)\rangle\langle s\hat{h}(\tau'_i)|B_i|s\hat{h}(\tau)\rangle\right) \tag{4.30}$$

$$= -J^2\varphi_i^s B_i^{++}B_j^{-+} + J^2\varphi_j^s B_j^{++}B_i^{-+} \sim O(J^2/\sqrt{N}). \tag{4.31}$$

where we have used that (from first order perturbation theory)

$$|\, s\hat{h}(\tau')\rangle = |\, s\hat{h}(\tau)\rangle + \sum_{m=1}^{k}\varphi_{i_m}^s|\, -s\hat{h}(\tau)\rangle + \ldots \tag{4.32}$$

where $\varphi_k^s = \frac{J\tau_k^z\langle -s\hat{h}(\tau)\,|\mathbf{a}_k\cdot\sigma|\,s\hat{h}(\tau)\rangle}{s|\mathbf{h}|} \sim O(1/\sqrt{N})$ and i_m indices the flipped l-bits in τ'. We have also introduced the short hand notation for the matrix elements of B_i with respect to the initial central spin field $\hat{h}(\tau)$,

$$B_i^{\pm\pm} = \langle \pm s\hat{h}\,|B_i|\,\pm s\hat{h}\rangle \tag{4.33}$$

As a result, the term corresponding to the interaction energy $\Delta^{(12)}$ in Eq. 4.21 predicts the occurence of resonances for $J \sim W/\sqrt{N}$, while for the term due to the matrix element deviation η these occur for $J \sim W/N^{3/4}$, since the difference in the intermediate state energies $\Delta E_{s\tau'_i} - \Delta E_{s\tau'_j} \sim O(W)$ due to the disorder terms Δ_i, Δ_j.

We now move on to the interference between the pair of channels with $\tilde{s} = -s$,

$$\Delta E_{s\tau'_{ij}} = \Delta E_{-s\tau'_i} + \Delta E_{-s\tau'_j} + 4s|\mathbf{h}| + \cdots \tag{4.34}$$

which provides a larger interaction energy as compared to Eq. 4.27:

$$\Delta^{(ij)} = 4s|\mathbf{h}| + \cdots = O(J\sqrt{N}) \tag{4.35}$$

The matrix elements are:

$$\mathcal{M} = \frac{J^2}{2}\left(\langle s\hat{h}(\tau'_{ij})|B_i| - s\hat{h}(\tau'_j)\rangle\langle -s\hat{h}(\tau'_j)|B_j|s\hat{h}(\tau)\rangle + \langle s\hat{h}(\tau'_{ij})|B_j| - s\hat{h}(\tau'_i)\rangle\langle -s\hat{h}(\tau'_i)|B_i|s\hat{h}(\tau)\rangle\right) \tag{4.36}$$

$$= J^2 B_i^{-+} B_j^{-+} + O(J^2/\sqrt{N}) \tag{4.37}$$

$$\mathcal{M}\eta = \frac{J^2}{2}\left(\langle s\hat{h}(\tau'_{ij})|B_i| - s\hat{h}(\tau'_j)\rangle\langle -s\hat{h}(\tau'_j)|B_j|s\hat{h}(\tau)\rangle - \langle s\hat{h}(\tau'_{ij})|B_j| - s\hat{h}(\tau'_i)\rangle\langle -s\hat{h}(\tau'_i)|B_i|s\hat{h}(\tau)\rangle\right) \tag{4.38}$$

$$= J^2\left(\varphi_i^s\left[B_j^{-+}B_i^{--} - B_j^{--}B_i^{-+} + B_j^{++}B_i^{-+} + B_j^{-+}B_i^{++}\right]\right. \tag{4.39}$$

$$\left.+\varphi_j^s\left[B_j^{-+}B_i^{--} - B_j^{-+}B_i^{++} - B_j^{++}B_i^{+-} - B_j^{--}B_i^{-+}\right]\right) \sim O(J^2/\sqrt{N}) \tag{4.40}$$

As in the previous case, the differences in energy in the intermediate state satisfy $\Delta E_{-s\tau'_i} - \Delta E_{-s\tau'_j} \sim O(W)$ due to the disorder terms Δ_i, Δ_j. As a result, the interaction energy $\Delta^{(12)}$ term in Eq. 4.21 predicts the occurence of resonances for $J \sim W/N^{5/6}$, while for the term from the matrix element deviation these occur for $J \sim W/N^{3/4}$.

Therefore, at $J \sim W/N \ll W/N^{5/6}$, the previous analysis shows that there are no second order resonances for final states with the central spin unflipped.

Central spin flipped ($s' = -s$)— In this case, rather than being paired by the intermediate state of the central spin $\tilde{s}$, it is convenient to pair the channels by whether the central spin flips simultaneously with the i'th l-bit or with the j'th. This way we can show that interference between channels does not change the onset of resonances at $J \sim W/N$. Again, for completeness, we list the four different amplitudes:

$$J^2\frac{\langle -s\hat{h}(\tau'_{ij})|B_i|s\hat{h}(\tau'_j)\rangle\langle s\hat{h}(\tau'_j)|B_j|s\hat{h}(\tau)\rangle}{(-\Delta E_{-s\tau'_{ij}})(-\Delta E_{s\tau'_j})} \tag{4.41}$$

$$J^2\frac{\langle -s\hat{h}(\tau'_{ij})|B_j|s\hat{h}(\tau'_i)\rangle\langle s\hat{h}(\tau'_i)|B_i|s\hat{h}(\tau)\rangle}{(-\Delta E_{-s\tau'_{ij}})(-\Delta E_{s\tau'_i})} \tag{4.42}$$

$$J^2\frac{\langle -s\hat{h}(\tau'_{ij})|B_i| - s\hat{h}(\tau'_j)\rangle\langle -s\hat{h}(\tau'_j)|B_j|s\hat{h}(\tau)\rangle}{(-\Delta E_{-s\tau'_{ij}})(-\Delta E_{-s\tau'_j})} \tag{4.43}$$

$$J^2\frac{\langle -s\hat{h}(\tau'_{ij})|B_j| - s\hat{h}(\tau'_i)\rangle\langle -s\hat{h}(\tau'_i)|B_i|s\hat{h}(\tau)\rangle}{(-\Delta E_{-s\tau'_{ij}})(-\Delta E_{-s\tau'_i})} \tag{4.44}$$

For each of the pairs:

$$\Delta E_{-s\tau''_{ij}} = \Delta E_{-s\tau'_i} + \Delta E_{s\tau'_j} + O(J) \tag{4.45}$$

$$\Delta E_{-s\tau''_{ij}} = \Delta E_{s\tau'_i} + \Delta E_{-s\tau'_j} + O(J). \tag{4.46}$$

Therefore, the interaction energy is only $\Delta^{(ij)} = O(J)$, which is again small for $J \sim W/N$ and does not lead to resonances at $J \sim W/N$.

However, unlike the previous case where the final state did not have the central spin flipped, the matrix elements for the four channels are quite different. For example, in the first pair (central spin flips with the i'th l-bit):

$$\mathcal{M} = \frac{J^2}{2}\left(\langle -s\hat{h}(\tau'_{ij})|B_i|s\hat{h}(\tau'_j)\rangle\langle s\hat{h}(\tau'_j)|B_j|s\hat{h}(\tau)\rangle + \langle -s\hat{h}(\tau'_{ij})|B_j| - s\hat{h}(\tau'_i)\rangle\langle -s\hat{h}(\tau'_i)|B_i|s\hat{h}(\tau)\rangle\right) \tag{4.47}$$

$$\sim \frac{J^2}{2}\left(B_i^{-+}B_j^{++} + B_i^{-+}B_j^{--}\right) \sim O(J^2) \tag{4.48}$$

$$\mathcal{M}\eta = \frac{J^2}{2}\left(\langle -s\hat{h}(\tau'_{ij})|B_i|s\hat{h}(\tau'_j)\rangle\langle s\hat{h}(\tau'_j)|B_j|s\hat{h}(\tau)\rangle - \langle -s\hat{h}(\tau'_{ij})|B_j| - s\hat{h}(\tau'_i)\rangle\langle -s\hat{h}(\tau'_i)|B_i|s\hat{h}(\tau)\rangle\right) \tag{4.49}$$

$$\sim \frac{J^2}{2}\left(B_i^{-+}B_j^{++} - B_i^{-+}B_j^{--}\right) \sim O(J^2) \tag{4.50}$$

This suggests that the amplitude in Eq. 4.21 can become of $O(1)$ because of the η term if $J \sim W/N$, since the intermediate state energy difference $\Delta E_{-s\tau'_i} - \Delta E_{s\tau'_j} \sim O(W)$, again due to the disorder terms. This is the expected result from the naive argument.

Let us check that no further cancellations arise from summing all four channels, focusing only on the η terms in Eq. 4.21 resulting from first pairing the amplitudes based on which l-bit the central spin flips together with. The resulting sum after the two pairings is:

$$\frac{J^2\frac{1}{2}\left[B_i^{-+}(B_j^{++} - B_j^{--})\right](\Delta E_{-s\tau'_i} - \Delta E_{s\tau'_j})}{\Delta E_{-s\tau'_{ij}}\Delta E_{-s\tau'_i}\Delta E_{s\tau'_j}} + \frac{J^2\frac{1}{2}\left[B_j^{-+}(B_i^{++} - B_i^{--})\right](\Delta E_{-s\tau'_j} - \Delta E_{s\tau'_i})}{\Delta E_{-s\tau'_{ij}}\Delta E_{-s\tau'_j}\Delta E_{s\tau'_i}} \tag{4.51}$$

Since the matrix element expressions in the square brackets are independent random variables of $O(1)$, the two terms in Eq. 4.51 are independent and cannot interfere destructively. Hence, we reproduce the estimate that $J \sim W/N$ produces $O(1)$ (second order) resonances, when the central spin flips.

This is physically sensible as delocalization in this model proceeds due to the interactions with the central spin, so that the divergent resonant subgraph should include classical configurations in which the central spin is flipped. We now analyze the resonant subgraph of second order processes. To recap, at $J \sim W/N$, the initial state $|s\hat{h}(\tau), \tau\rangle$ is resonant with O(1) neighbours of the form $|-s\hat{h}(\tau'_{ij}), \tau'_{ij}\rangle$ ('step 1 second order' resonances). Now consider the $\sim N^2$ states accessible from a step 1 resonant state with two more l-bits flipped and the central spin returned to s (as the central spin flips at each step). In order for the l-bit pair kl to be resonant, the state must have energy $|\Delta E_{s\tau'_{ijkl}}| < \frac{J^2}{W}$ (again from second order perturbation theory). As $\Delta E_{s\tau'_{ijkl}}$ does not contain the energy change on flipping the central spin of order $\sim J\sqrt{N}$ and this value is much greater than the level spacing of order W/N^2 of $\Delta E_{s\tau'_{ijkl}}$ over pairs kl, the pairs kl that satisfy the resonance condition at step 2 are typically not the same as the pairs ij that satisfy the resonance condition at step 1. Thus, the resonant subgraph includes new l-bit flips at step 2 as compared to step 1.

At step 3, the $\sim N^2$ accessible states involve three l-bit pair flip-flops with the central spin again flipped to $-s$. The resonance condition $|\Delta E_{-s\tau'_{ijklmn}}| < \frac{J^2}{W}$, which is analogous to step 1, except for the flipped l-bits $ijkl$, picks *different* pairs as compared to step 1, because the energy difference,

$$\Delta E_{-s\tau'_{ijklmn}} = \Delta E_{s\tau'_{ijkl}} - 2s|\mathbf{h}| - 2\tau^z_m(\Delta_m - sJ\mathbf{a}_m \cdot \hat{h}) - 2\tau^z_n(\Delta_n - sJ\mathbf{a}_n \cdot \hat{h}) \tag{4.52}$$

$$+ 4Js(\tau^z_i \mathbf{a}_i \cdot \hat{h} + \tau^z_j \mathbf{a}_j \cdot \hat{h} + \tau^z_k \mathbf{a}_k \cdot \hat{h} + \tau^z_l \mathbf{a}_l \cdot \hat{h}) + O(J^2/\sqrt{N}) \tag{4.53}$$

contains new terms of $O(J)$ when compared with $\Delta E_{-s\tau'_{mn}}$, e.g. $4Js\tau^z_i \mathbf{a}_i \cdot \hat{h}$, due to the flipped l-bits $ijkl$. Since this is much larger than the level spacing W/N^2 of varying $\Delta E_{-s\tau'_{ijklmn}}$ over pairs mn, different pairs of resonant l-bits will be picked compared to step 1. Moreover, the interaction terms of order $\sim J^2/|\mathbf{h}| \sim W/N^{3/2}$, not explicitly written out in Eq. 4.52 for simplicity, ultimately determine which pair mn is resonant, since similarly to the $O(J)$ terms, $W/N^{3/2}$ is much larger than the level spacing W/N^2. Thus, the pairs mn that satisfy the resonance condition $|\Delta E_{-s\tau'_{ijklmn}}| < J^2/W$ are not the same as the pairs ij that satisfy the resonance condition $|\Delta E_{-s\tau'_{ij}}| < J^2/W$ at step 1 as the new terms due to having flipped other l-bits have significantly changed it. When $J \sim W/N$, we therefore find $O(1)$ *completely new* resonant states among the configurations accessible by flip-flops at *each* step.

From the above construction, we see that the resonant subgraph of the hypercube generated by second order processes is tree-like for small enough J, as at each step a completely different pair is resonant (see Fig. 4.13). The expected branching number of the tree, i.e. the number of resonant l-bits at each step, is $K \sim J^2N^2/W^2$. This random trees undergo a continuous percolation transition at $K = K_c \sim 1$ at which the probability

that the resonant subgraph is infinite starts growing from zero. However, when the expected branching number K becomes too large, we expect the tree approximation to break down as short loops in the interaction graph become very important, i.e. when it becomes likely that a certain flipped l-bit will be flipped back to the original state after a few moves. As the critical branching number K_c is of order one, our treatment is self-consistent within the MBL phase. This provides a $J_c(N) \sim \sqrt{K_c}W/N$ beyond which second order resonances guarantee that typical eigenstates cannot remain localized on the hypercube.

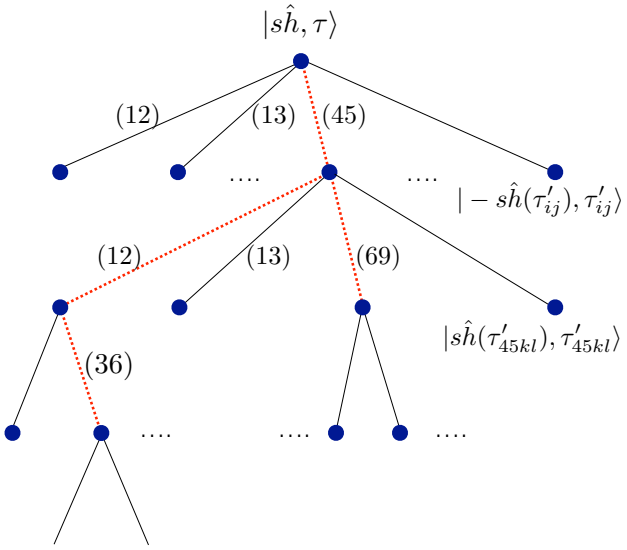

Figure 4.13: Schematic of graph generated by second order processes on the hypercube of classical configurations. At each step, there are $Z \sim N^2$ pairs that may be flipped (edges) of which only $K \sim O(1)$ are resonant (red dashed). The shown resonant subgraph has $n = 5$ states.

4.5.4 Properties of resonant subgraphs

In this section, we model the resonant subgraphs generated by second order processes by independent bond percolation on the hypercube with second neighbor bonds placed with probability $p \sim J^2/W^2$ such that the expected number of resonant bonds at each step is preserved. The following properties of resonant subgraphs then follow from percolation theory [96, 97]:

1. For $J < J_c$, the number of resonant subgraphs of size n per site in Hilbert space is given by:

$$c(n) \propto n^{-5/2} e^{-n/n_\xi} \tag{4.54}$$

with

$$n_\xi \sim \left(\frac{J - J_c}{J_c} \right)^{-2} \tag{4.55}$$

n_ξ is the characteristic number of sites in a resonant subgraph in the localized phase, which diverges at the transition. At the transition and for $n \ll n_\xi$ in the vicinity of the transition, the subgraphs are 'critical' and are distributed according to a power-law. A randomly chosen site in the hypercube lies in a connected cluster of size n with probability

$$p(n) \propto nc(n) \sim n^{-3/2} e^{-n/n_\xi} \tag{4.56}$$

2. At the critical point $J = J_c$ at finite N, we expect Eq. 4.56 to describe the statistics of cluster sizes with a cutoff size n_ξ that diverges with increasing N. Above the upper critical dimension for percolation, the largest cluster scales with the volume of space to the 2/3 power,

$$n_\xi \sim V^{2/3} \sim 2^{\frac{2}{3}N} \tag{4.57}$$

Thus, at the critical point at size N, the largest resonant subgraphs, while exponentially large in N are still an exponentially small fraction of the hypercube. We expect that such large subgraphs are equally likely to include states in which l-bit i is up or down.

3. For $J > J_c$, there is a unique giant component which absorbs a fraction $P \propto |J - J_c|^1$ of the total volume 2^N of the hypercube. The remaining clusters are of finite size with a distribution $\propto n^{-3/2} e^{-n/n_\xi}$. In the giant component, all l-bits are equally likely to be up or down, while the remaining clusters have only a finite number of flipped l-bits (maybe with the exception of a few rare ones of size $\geq O(N)$, c.f. point 2 below as well).

4.5.5 Consequences

We now assume that the actual many-body eigenstates are simply delocalized over the resonant subgraphs generated by second order processes. This leads to several quantitative predictions about the localized phase, which agree with numerical observations in

Section 4.4. On the other hand, the predictions for the crossover region do not explain the numerical data very well and suggest that higher order processes are important.

1. The transition from localized to delocalized takes place at $J = J_c(N) \sim W/N$. This is consistent with our numerical observations in Section 4.4. However, as we discuss below, higher order processes may decrease J_c logarithmically in N.

2. In a cluster with m edges, the number of flipped l-bits is at most $2m$ (as at most two distinct l-bits can flip on every edge of the cluster, see Fig. 4.13). Thus, the entanglement entropy of only $O(m)$ l-bits can be non-zero in a many-body eigenstate delocalized over this cluster. In the localized phase, the clusters are finite and thus

$$[S_1] = O(1/N) \tag{4.58}$$

   This explains the remarkable collapse in Fig. 4.2 in the localized phase. Indeed, the entire distribution of S_1^i has weight only at zero as $N \to \infty$ which implies that $\Delta S_1 \sim 1/N$ in the localized phase.

3. Within this model, the participation ratio PR of an eigenstate $|\, E\rangle$ is simply given by the size n of the associated cluster. Thus, the PR distribution is given by Eq. 4.56. This explains several aspects of the numerical PR distributions in Fig. 4.9: (1) the PR distribution is nearly N-independent for small JN and has a sharp cut-off n_ξ, (2) the cut-off scales exponentially with N as predicted by Eq. 4.57 in the crossover region. However, the power law regimes suggested by this model are not clearly seen in the numerics.

4. At finite N, the localized phase in the vicinity of the transition is predicted to be heterogenous across eigenstates, l-bits and samples as each initial configuration τ provides a different landscape for building the resonant subgraph. However, this heterogeneity vanishes as $N \to \infty$ as almost all the l-bits are unentangled in eigenstates (see the discussion of $[S_1]$ above), so this does not explain the numerical observations in the crossover region.

5. Another possible model of the crossover region is that it is described by the bond percolation problem at $J > J_c$ with a giant component. For eigenstates delocalized on the giant component, $S_1^i = 1$, while for eigenstates delocalized on finite clusters, S_1^i is non-zero only for a finite number of l-bits. This model therefore predicts that the distribution for S_1 becomes bimodal as $N \to \infty$ with $p(S_1 = 0) = (1 - P)$ and $p(S_1 = 1) = P$, where $P \sim |J - J_c|$ is the fraction of sites belonging to the giant component.

However, the numerics presented in Section 4.4 do not support this picture. First, there is no evidence of a giant component in the PR distributions in the crossover region in Fig. 4.9(b). Furthermore, this percolation model does not predict a significant variation of S_1^i between l-bits i in the same state. Numerically, this is the biggest source of the variation in S_1^i in the crossover region (see Fig. 11).

4.5.6 Higher order processes

The Anderson localization transition for a single particle hopping on a Bethe lattice [107] with large coordination Z takes places at hopping strength $J \sim W/Z \ln Z$, where W is the bandwidth of the independently sampled disorder on each site [99]. Altshuler, *et al.* argued that this result applies directly to the 'hopping' problem in Fock space induced by two-body interactions in a disordered quantum dot [91]. They further conjectured a phase diagram in which the eigenstates are localized for $J < W/Z \ln Z$, delocalized sparsely in Fock space for $W/Z \ln Z < J < W/Z$ and fully delocalized for $W/Z < J$.

Physically, the transition at $W/Z \ln Z$ arises due to long distance, high order resonances in the Fock space. The associated states appear mixed from the point of view of few body observables, just as the GHZ state does, an equal superposition of all spins either up or down, even though their participation ratios remain a vanishingly small fraction of the total volume of the hypercube. They further would not exhibit level repulsion as states neighboring in energy typically do not overlap.

There are a number of assumptions that go into mapping the interacting dot model onto the Bethe lattice localization problem. The most crucial of these are 1) that interference due to the presence of loops on the hypercube is unimportant at long distances and 2) that the energy correlations are likewise unimportant. These are plausible but not rigorously justified. Indeed, even in the Bethe lattice problem, whether this regime is a phase or a slow crossover is still debated [102, 103, 104].

The central spin model can be mapped to the Bethe lattice problem under similar assumptions as the quantum dot model of Altshuler, *et al.* with $Z = N$. This mapping predicts that higher order processes destabilize MBL as $N \to \infty$ at asymptotically smaller J than the second order processes described in Section 4.5.3. If this is true, then it seems that there are two possibilities:

1. There is a direct transition from the localized to the ETH phase at $J_c \sim W/N \ln N$. If this is the case, then it must be that at the small sizes N that we can reach numerically, whatever non-perturbative processes cause the sparsely delocalized states

to fully thermalize in the limit of large N are ineffective at these small N, and we primarily observe behavior governed by the second order picture up to the scale $J_c \sim W/N$. This is why we develop a detailed picture of the second order percolation model in Section 4.5.5.

2. There is an intermediate delocalized ergodic phase between a transition at $J_c \sim W/N \ln N$ and another transition on the scale $J^* \sim W/N$. This would imply, for example, that $[S_1]$ transitions at asymptotically smaller coupling than $[r]$ by this factor of $\ln N$.

We have attempted to interpret our numerical results from both of these view points in Section 4.4, but the results are inconclusive, so we leave this difficult question about the possibility of such an intermediate phase in our central spin model for future work.

4.6 Conclusions

We have explored a central spin model for many-body localization and thermalization. We show that a single central spin can serve as a 'seed' for interactions between otherwise noninteracting l-bits and thus produce a thermal bath. From the point of view of finite dimensional disordered systems, our model captures the interaction between a small inclusion and the layer of its immediate neighbors. As the number N of immediate neighbors grows with dimensionality d, the coupling necessary for this layer to thermalize tends to zero. This bath, 'nucleated' by just one central spin, can then thermalize further layers following a 'layer-by-layer' version of the arguments presented in Ref. [85] and accordingly destabilize MBL.

We show, based both on numerics for small systems and on perturbative arguments, that a central-spin system with N l-bits exhibits an MBL phase and phase transition in the limit of large N if we scale the interaction down with N. This dynamical transition occurs at an interaction strength that is thermodynamically insignificant in the large-N limit. We argue that the MBL phase can be understood as localization on a hypercube of Fock states [91]. We discuss and explore the possibility of an intermediate "delocalized but nonergodic" phase between the MBL and thermal phases. Our numerical results on small systems do not allow us to draw a conclusion about whether or not such an intermediate phase exists in this model. Our results also apply to more realistic models of the l-bits that include diagonal interactions between them; with diagonal interactions between the l-bits, first order processes can already produce a divergent sub-graph of resonant configurations for $J > J_c(N)$.

We have also examined the finite-size behavior near the phase transition, particularly of the single-l-bit entanglement within eigenstates. We find that the probability distribution of this quantity is bimodal near the transition, with peaks near full entanglement and near zero entanglement. This heterogeneity is shown to occur both across l-bits within single eigenstates as well as across eigenstates for a single l-bit. This is a feature of the finite size data that has also been seen in 1d models and is not yet well understood [86, 87]. The finite size data for our central spin model do not scale well, indicating that the sizes we can access numerically are well short of any asymptotic large-N scaling regime. One indication of this is that the sample-to-sample variations are small, but appear to be accelerating in their growth with increasing N.

Let us speculate on possible scenarios for the phase transition out of the localized phase in to a delocalized phase (which may be non-ergodic or obey ETH). In the limit of large systems, each sample s has its own transition point $J_c(s)$. Due to the quenched disorder, we expect the probability distribution of these transition points to have a width $\Delta J_c \sim J_c/\sqrt{N}$ [92, 93]. Within a single finite-size sample the transition may be sharper than this, of similar width, or broader than this. The apparent accelerating increase of the sample-to-sample differences seen numerically (Fig. 4.8(d)) argues against the last of these possibilities. This leaves the two other possibilities. At the accessible system sizes, the inter-sample and within-sample variations are comparable; if this persists to large N, we would expect the rounding of the transition to be of width $\sim J_c/\sqrt{N}$ both within samples and between samples. The decreasing trend of the peak with N in Fig. 4.8(c) suggests that the inhomogeneity between eigenstates in the l-bit-averaged entanglement at the transition might go away at large N, thus making all eigenstates in one sample similar in this respect (although it seems likely they will still differ in which l-bits are less versus more entangled).

The other possible scenario is that the transition in a single sample is sharper by a power of N than the variation in $J_c(s)$ between samples. A limiting case of this scenario in equilibrium is provided by first order transitions with quenched randomness [105, 106]. In this scenario, if we scale J by the sample averaged J_p, then at J_p, almost all samples will be away from their transition and be in their MBL or delocalized phases respectively. Almost all the variation in S_1 then comes from sample-to-sample variations. We are clearly not in this regime for the sample sizes we study numerically here, but it remains possible that this happens at much larger N. To study the single-sample transition in this regime, one should instead scale J by the sample-specific transition point $J_c(s)$. It is an interesting question to ask how sharp this transition could possibly be. What would be the nature of the sharpest possible, thus most discontinuous, MBL-to-delocalized phase transition? We leave these questions for future research.

Chapter 5

Discussion and Outlook

In Chapters 2 and 3 we showed numerical evidence that ergodic systems generically heat-up to infinite temperature with periodic driving, whereas many-body localized systems can remain stable as an extensive set of localized conserved quantities survives the driving. These results, which additionally have been supported by rigorous theoretical arguments [108] and experiments in cold-atom systems [109], laid out the foundations for a series of interesting and also unexpected developments. Because individual MBL eigenstates do not satisfy the ETH, they can host a multitude of different types of order such as symmetry breaking that are forbidden in equilibrium due to thermal fluctuations [110]. The possibility of MBL phases to host different types of order combined with their stability with respect to periodic driving has resulted in the discovery of new types of ordered phases [111, 112] – impossible to realize in the static setting – with the most notable example of the Floquet time crystal exhibiting time translation symmetry-breaking, which has already been observed experimentally [113, 114]. On the other hand, even though thermalization of driven ergodic systems implies their collapse to a featureless infinite temperature state at long times, the transient dynamics turns out to be more interesting. In fact, the latter can be approximated by the dynamics of an effective local Floquet Hamiltonian obtained from the Magnus expansion truncated to a few terms [115, 116, 117, 118, 119]. This regime known as pre-thermalization lasts until heating effects take place at a time scale which increases exponentially with the frequency of the drive. Theoretical bounds for the optimal truncation of the Magnus expansion exist, but have not been verified numerically yet.

In Chapter 4 we showed that a two-level thermal inclusion interacting with N localized bits subject to random fields leads to their thermalization when the interaction strength is larger than a critical value that decreases as $1/N$. Our results indicate that the MBL

phase is likely unstable in systems with short-range interactions and quenched randomness in dimensions d that are high but finite. Moreover, the patterns of single bit entanglement at the transition vary significantly between eigenstates of the same sample, which provides evidence for a heterogeneous discontinuous transition out of the localized phase with respect to single-site observables. A complete microscopic theory for the nature of the MBL-to-thermal transition is still missing. However, these results suggest that a complete understanding of the effects of thermal inclusions will constitute an important step forward [120]. To this effect, machine learning techniques, the topic of the second part of this thesis, might become important tools towards the general identification of relevant order parameters for this phase transition, as well as towards a data-driven algorithm to identify thermal inclusions (c.f. discussion in Chapter 8).

As we have seen in the first part of this thesis, the physics of closed quantum systems away from equilibrium constitutes rich territory and many important questions remain unanswered despite remarkable progress over the last decade. Addressing these questions is now more urgent than ever with the increasing sophistication of experimental techniques involved in the simulation and manipulation of isolated synthetic quantum matter, e.g. in cold-atom systems.

Part II

Machine-learning phases of matter

Chapter 6

Introduction

6.1 Machine learning

Machines that match humans in general intelligence and that are able to process diverse types of information across different domains have been expected since the invention of computers [121]. Indeed, the development of artificial intelligence (AI) included significant challenges manifest in several periods of reduced funding and general interest in the field known as the AI winters. However, today, AI is a thriving field in science and technology with many applications. These include human level speech [122] and image recognition [123], the ability to make diagnostics in medicine [124], and contributions to basic scientific research across different disciplines [125, 126].

One of the earliest approaches to AI was based on expert systems [121]. These required a knowledge-base of facts painstakingly hand-coded from human knowledge from which inferences about new inputs were made. For example, in a task of predicting hand-written digits in digitized images, one would handcraft rules or heuristics for distinguishing the digits based on the shapes of the strokes, which would result in a proliferation of rules and of exceptions to the rules and so on. These systems were expensive to develop, validate and maintain, and the interest in them eventually waned.

On the other hand, the arguably most successful approach to AI has been machine learning [127, 128]. This category refers to the set of computer algorithms that given a specific task acquire their own knowledge by extracting the relevant patterns from raw data. And then use these patterns to perform decisions given new data that has not been observed before. Machine learning algorithms can be divided into three types: supervised

learning, unsupervised learning and reinforcement learning, which are distinguished by the nature of the task they are designed to accomplish.

Supervised learning

In supervised learning, the goal is to learn to map inputs $\boldsymbol{x}$ to outputs y. The machine learns to perform this mapping through a training set of examples, i.e. a list of input and associated output pairs $\mathcal{T} = \{\boldsymbol{x}_i, y_i\}_{i=1}^{N_\mathcal{T}}$. Each input can be generically shaped as vector $\boldsymbol{x}_i$, where each component is called a *feature*. For example, if $\boldsymbol{x}_i$ describes a specific image, each feature can correspond to the value of a pixel in the image.

Supervised learning is further discriminated based on the nature of the output y. If y is a real variable, the problem is known as regression. On the other hand, if y is categorical, i.e. it takes only c possible values $\{0, 1, 2, \ldots, c-1\}$ corresponding to c different categories, then the problem is known as classification.

Unsupervised learning

In unsupervised learning, the algorithm only has access to a set of training inputs $\{\boldsymbol{x}_i\}$. The distinguishing feature with respect to supervised learning is the absence of labels. Different tasks fall under this category. In density estimation, one tries to learn the data-generating probability distribution associated with $\{\boldsymbol{x}_i\}$. Other tasks involve finding certain structures in the data. For example, clustering algorithms search for optimal partitions of the dataset into groups whose elements have similar properties (c.f. K-means algorithm), while manifold learning searches for a lower dimensional but efficient representation of the data points $\boldsymbol{x}_i$ (e.g. Principal Component Analysis).

Reinforcement learning

In reinforcement learning [129], an agent interacts with an environment, and at a given time can take a set of possible actions. The goal is to learn how to map the state of the environment to an action, in order to maximize future rewards, which might come only many steps later. The agent is not given a priori any information about which actions to take, but must discover which actions yield the most reward by exploring and trying them out.

6.1.1 Example: Logistic regression

In Chapter 7 of this thesis we will be classifying phases of matter, so in preparation we review a simple machine learning algorithm for binary classification called logistic regression and use this example to introduce several important machine learning concepts.

In logistic regression the probability of an input $\boldsymbol{x}$ to be classified as say class 1 is given by $p(y = 1|\boldsymbol{x}; \boldsymbol{w}) = \sigma(\boldsymbol{w} \cdot \boldsymbol{x} + b)$. The function σ is the sigmoid function $\sigma(z) = \frac{1}{1+e^{-z}}$ and squashes its output to the interval $[0, 1]$ so that it describes a probability. Due to normalization, the probability of the input to be classified as class 2 is $p(y = 2|\boldsymbol{x}; \boldsymbol{w}, b) = 1 - p(y = 1|\boldsymbol{x}; \boldsymbol{w}, b)$. The parameters $\boldsymbol{w}$ and b defining the model are called weights and bias, respectively. Notice that even though the argument of the sigmoid $\boldsymbol{w} \cdot \boldsymbol{x} + b$ is linear in $\boldsymbol{x}$, it can be non-linear in an original set of features $\tilde{\boldsymbol{x}}$ by including in the vector $\boldsymbol{x}$ e.g. polynomial features, such as $\tilde{x}_1\tilde{x}_2$, or other non-linear functions of the original features.

The learning algorithm determines the best parameters $\boldsymbol{w}, b$ by optimizing a performance metric in the training set. For classification, the common metric is the maximum likelihood estimator. This means the optimal parameters are those that maximize the probability of observing the input, output pairs in the training set, i.e. $\prod_{i=1}^{N_\mathcal{T}} p(y = y_i|\boldsymbol{x}_i; \boldsymbol{w}, b)$. It's standard, however, to equivalently reframe the problem as minimizing the negative log-likelihood:

$$J(\boldsymbol{w}, b) = -\frac{1}{N_\mathcal{T}} \sum_{i=1}^{N_\mathcal{T}} \ln p(y = y_i|\boldsymbol{x}_i; \boldsymbol{w}, b) \tag{6.1}$$

The solution $\boldsymbol{w}, b$ to the minimization of the cost function $J(\boldsymbol{w}, b)$ is found using a gradient descent algorithm, as there is no closed form solution [128]. The process of optimizing the cost function over the training set is known as the training phase of the algorithm.

To evaluate the quality of the algorithm, instead of analyzing the negative log-likelihood corresponding to the optimal parameters, it is common to focus on the accuracy of the predictions of the model as it constitutes a more intuitive metric. This assumes the number of examples in each of the classes is balanced, otherwise alternative metrics might be more appropriate. If after training, the model captured enough knowledge to distinguish the two classes on the training set, then we expect its accuracy to be high.

However, the central challenge in machine learning is to find an algorithm that performs well on data that it has not seen before; this ability is known as the generalization. Thus, ultimately, the quality of a classification algorithm is determined by finding the accuracy on

a test set formed by different samples from the ones used in training: $\mathcal{T}' = \{\boldsymbol{x}'_i, y'_i\}_{i=1}^{N_{\mathcal{T}'}}$. One important factor that contributes to the discrepancy between the training set accuracy and the test set accuracy is if after training, the model learns details that are not generalizable patterns of the data.

To address this problem, it is common to employ regularization techniques. These techniques work by favouring certain solutions over others, such that the unpreferred solution is only selected if it fits the training data significantly better than the preferred solution.

One common regularization technique is weight decay. In logistic regression, weight decay penalizes solutions with larger weights $\boldsymbol{w}$. More concretely, the cost function becomes:

$$J(\boldsymbol{w}, b) = -\frac{1}{N_{\mathcal{T}}} \sum_{i=1}^{N_{\mathcal{T}}} \ln p(y = y_i | \boldsymbol{x}_i; \boldsymbol{w}, b) + \gamma ||\boldsymbol{w}||^2, \tag{6.2}$$

where γ defines an hyperparameter. This is a quantity that as opposed to the weights and bias, is not learned from the training set, but is chosen so that it results in the optimal performance in a test set. At very small γ, the weights are not significantly constrained, so the model that learns the most about the training set by minimizing the negative log-likelihood is selected. On the other hand, at large γ, the optimal solution is a model with small weights, and it might be too simple to capture all the details that are also relevant to make predictions on new data. So typically, the best generalization performance occurs for some intermediate value of γ.

The previous discussion on the role of the regularization parameter is related to a more general idea in machine learning concerning the generalization error known as bias-variance trade-off [128]. There's usually two sources contributing to the generalization error. The bias error is related to the inability of the model to learn all the complexity associated with the learning task. When the bias is large, the model doesn't perform well either on the training or the test set. This regime is known as underfitting. When the model complexity is very large, the bias is reduced, but now the model is susceptible to capture noise or unrepresentative properties of the training data. Consequently, predictions on new data are volatile with respect to details in the training set, which results in bad generalization. This is the overfitting regime. Therefore, the general goal is to aim for a model whose complexity corresponds to the optimal balance between the two types of errors.

6.2 Learning phase transitions

Over the past year there has been a surge of research activity applying machine learning to several quantum physics problems. However, the relationship between the two fields has

been a symbiotic one given the promise of quantum computing to improve the performance of machine learning algorithms [130].

As mentioned above, the applications of machine learning to quantum physics have been varied [131]. These include learning thermal and quantum phase transitions with supervised and unsupervised learning, including the MBL-to-thermal transition, representing physical wavefunctions with machine learning models and designing algorithms combining machine learning and the renormalization group.

Most of the works so far have been based on a class of algorithms known as artificial neural networks, whose basic units are called neurons. These neural network algorithms and, in particular, deep neural networks whose architecture is formed by several layers of neurons were responsible for recent breakthroughs in computer vision, speech recognition, and natural language processing tasks with significant technological impact. It is believed that underlying the success of deep learning is the ability to develop a hierarchy of concepts, from basic ones at the bottom layers closer to the input to more complex ones at the top layers; this is particularly clear in image recognition tasks [132]. However, that deep architectures provide practical advantages compared to shallow architectures when working with data originating from physical systems is still open to debate [133].

Going back to the task of learning phases of matter, which is the subject of Chapter 7, the pioneering work in Ref. [134] showed that neural networks could perform the supervised learning task of associating spin configurations from the 2D Ising Model with the corresponding phase and even extract a certain critical exponent. The input to the learning algorithm were configurations of the system sampled from the equilibrium Gibbs distribution at different temperatures, obtained from Monte Carlo simulations. Each configuration was further labelled by the corresponding phase, depending on whether the temperature of the corresponding Gibbs distribution was below or above the transition temperature separating a ferromagnetic from a paramagnetic phase. The same work showed that neural networks were also able to discriminate topological phases, however, in this case, successful learning was only possible with convolutional neural networks (CNN), whose architecture directly enforces locality and translational invariance. Indeed, the CNN was able to learn that the topological ground states of the model under study were characterized by an extensive number of local constraints.

There have also been several unsupervised learning approaches to discriminating phases of matter that do not require the precise knowledge of the boundaries between different phases. Refs. [135, 136, 137] showed that for different models the Principal Component Analysis (PCA) algorithm, which takes configurations of the system as input without any label, was able to reveal order parameters that describe the phase transition. However, it only works if the order parameter is related to a linear function of the input to the

algorithm, otherwise the success of PCA depended on further feature engineering. For this reason, there have been attempts to go beyond PCA using autoencoders [138]. Ref. [139] introduced a confusion scheme to detect the phase transition when a certain parameter is varied by doing supervised learning assuming different values for the critical parameter. Under the reasonable assumption that maximum discrimination between the two sets of configurations should occur when the guess for the critical parameter is close to the exact one, this provides a way to detect the transition. Refs. [141, 140] also used a similar approach. In the context of many-body localization, where the knowledge of the precise boundaries to thermal phases are not known precisely for any model, a mixed approach was considered, where the data obtained when the Hamiltonian is deep in each of the phases is labelled, but the data originating from parameters near the crossover region is not labelled. The associated cost function optimized a neural network model to classify correctly the labelled data, and in addition to be confident about the predictions on the unlabelled data [142]; see also [143].

In the next chapter, we introduce another supervised learning algorithm, Support Vector Machines, for learning phase transitions, which allows to obtain interpretable non-linear order parameters associated with a phase transition, thus overcoming a general difficult task for neural network models.

Chapter 7

Kernel methods for interpretable machine learning of order parameters

7.1 Introduction

From a physics perspective, the problem of searching for patterns in experimental data has long driven theoretical progress which, by definition, relates the patterns to underlying postulates of our physical theories. Hence, in the modern push to automate the discovery of important and novel physical features in data [144, 134, 135, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 139, 156, 157, 154, 158, 159, 160, 161, 162, 163, 138, 164, 136, 142], physicists must be mindful of the interpretability of machine learning results if they are truly meant to drive the theoretical process.

In condensed matter, physicists face the ultimate big-data challenge. One must search practical measurements, obtained from the exponentially-large state space of a system or model, for patterns which relate to underlying theoretical paradigms. The recent success of neural networks in classifying phases of matter [134, 147, 156, 146, 145, 148, 142] has been encouraging, in that it demonstrates how relatively standard supervised learning tools can be repurposed for calculations in condensed matter physics. However, contrary to industry applications of machine learning, where performance is the prime metric of success, in physics it is generally desirable to further tie the outcome to some theoretical structure, which can eventually be used e.g. to make predictions. Neural network behavior can indeed be interpreted on the simplest models of statistical mechanics, such as the demonstration in Ref. [134] that the magnetization order parameter of the two-dimensional (2d) Ising model is learned by the weights of the hidden units. However, in general for more

complicated models, similar success in relating network structure to non-linear or non-local order parameters is challenging, especially in the case of deep neural networks. This lack of interpretability presents a challenge for the goal of driving theoretical progress with machine learning. It is therefore crucial that the condensed matter community survey the performance of interpretable machine learning algorithms on data obtained from models of interest to condensed matter physics.

In this chapter we study interpretable supervised learning algorithms applied to the discrimination of phases of matter in large synthetic data sets produced by numerical simulation. We focus on a particular class of machine learning algorithms called support vector machines (SVMs). In particular, we ask whether SVMs in the supervised learning setting are able to discover the mathematical structure of physical order parameters. We first introduce the SVM algorithm and describe its general properties, with particular focus on the *kernel trick*. This trick allows us to perform linear regression on non-linear features of the data, without explicitly generating the features. In particular, this is relevant when applied to Monte Carlo data from many-body systems, which typically have high-dimensional physical states. Without the kernel trick, it would be unpractical to generate a set of non-linear features, which could be exponentially large. In Section 7.3, we perform phase classification on Monte Carlo configurations produced from the 2d Ising model, a conserved-order-parameter Ising model (COP) [135], and the 2d Ising gauge theory. We find that SVMs with a quadratic polynomial kernel can discriminate the phases of the 2d Ising model by learning the correct physical order parameter (the squared magnetization per spin). In the case of COP model, we find that SVMs are able to discover the non-trivial order parameter, devised by Wang [135] through visualization of dimensionally-reduced data, with less human intervention. Finally, we show that the 2d Ising gauge theory, whose $T = 0$ ground state is defined by local plaquette constraints but no conventional order parameter, can be discriminated by SVMs with polynomial kernels. In this case, no polynomial with order less than four can identify the ground state, indicating that the SVM is able to learn the original Hamiltonian in order to evaluate whether the local constraints on four Ising degrees of freedom are satisfied for each plaquette.

7.2 Support Vector Machines

Linear support vector machines [128] were initially designed to perform binary classification. They belong to the class of supervised learning algorithms that, given an input $\boldsymbol{x} \in \mathbb{R}^p$, predict the class $y \in \{-1, 1\}$ in which it is most likely to belong. In the context of this chapter, the input $\boldsymbol{x}$ will represent a spin configuration from a statistical mechanical model of interest in condensed matter physics, in which the different components of

$\boldsymbol{x}$, called *features* in the machine learning literature, correspond to the spins at different lattice sites. The output y will label the thermodynamic phase it was sampled from. The main idea behind SVMs is to find the hyperplane, defined by $\boldsymbol{w} \cdot \boldsymbol{x} + b = 0$, that best separates the two classes. Formally, this is achieved by solving the optimization problem:

$$\begin{aligned} &\arg\min_{\boldsymbol{w},\xi_i,b} \left\{ \frac{1}{2}\boldsymbol{w}\cdot\boldsymbol{w} + \frac{1}{\gamma}\sum_{i=1}^{N}\xi_i \right\} \\ &\text{such that } y^{(i)}(\boldsymbol{w}\cdot\boldsymbol{x}^{(i)} + b) \geq 1 - \xi_i,\ i = 1 \dots N. \end{aligned} \tag{7.1}$$

Here, γ is a constant, ξ are "slack" variables (described with more detail below) and N is the number of input samples, called the training set. Finding optimal parameters $\boldsymbol{w}$ and b, which can then be used to make predictions on a test set, is the goal of supervised learning with SVMs. Unlike say a feed-forward neural network [134], this model does not provide probabilistic predictions. However, the optimization problem Eq. 7.1 is equivalent to $\arg\min_{\boldsymbol{w},b} \left\{ \frac{1}{2}\boldsymbol{w}\cdot\boldsymbol{w} + \frac{1}{\gamma}\sum_{i=1}^{N} \max\left(0, 1 - y_i d(\boldsymbol{x}^{(i)})\right) \right\}$ with $d(\boldsymbol{x}^{(i)}) = \boldsymbol{w}\cdot\boldsymbol{x}^{(i)} + b$. Because the hinge function $\max(0, 1 - yd)$ approximates the misclassification error $\Theta(-yd)$ [127], this can be viewed as an approximation to minimizing the misclassification error with a so-called ℓ_2-norm regularization (which penalizes unnecessary coefficients of $\boldsymbol{w}$). This analogy also explicitly shows that γ can be interpreted as a regularization parameter.

When the ξ_i variables are constrained to be zero in Eq. 7.1, this corresponds to the hard margin case and the optimization algorithm has a solution only for linearly separable classes. In this case, the solution corresponds to the hyperplane $\boldsymbol{w}\cdot\boldsymbol{x} + b = 0$ for which the *margin*, defined as the minimum distance $d_{\min} = 1/||\boldsymbol{w}||$ of the data samples to the hyperplane, is maximum. Thus, this algorithm finds the hyperplane which maximizes the training set margin and provides the most confident predictions on new inputs. On the other hand, perfectly separable data is not a typical property of datasets and in general the constraints $y^{(i)}(\boldsymbol{w}\cdot\boldsymbol{x}^{(i)} + b) \geq 1$ are not feasible. Hence, it is necessary to introduce the slack variables ξ which allow a training set input $\boldsymbol{x}^{(i)}$ to violate the margin at a cost of ξ_i/γ. For a given training sample, the slack variable ξ_i can take different values. $\xi_i = 0$ if the margin is not violated; $0 < \xi_i \leq 1$ if it is on the correct side of the hyperplane but violates the margin; and $\xi_i > 1$ if the sample is misclassified. In general it is necessary to use a test set to find the optimal parameter γ. This will provide the best trade-off between minimizing training errors and the model complexity.

After optimization (i.e. training), the class to which a new input $\boldsymbol{x}$ belongs is predicted as $y = \text{sign}(d(\boldsymbol{x}))$, hence $d(\boldsymbol{x})$ is referred to as the *decision function* in the machine learning literature. Even though the previous optimization problem can be solved by quadratic programming [128], it also admits a dual formulation whereby the primal variables $\boldsymbol{w}$, ξ, b

are eliminated and the optimization is performed over N dual variables α_i, which are the Lagrange multipliers associated with each constraint in Eq. 7.1. The optimal parameter $\boldsymbol{w}$ is then expressed as $\boldsymbol{w} = \sum_{i=1}^{N} \alpha_i y_i \boldsymbol{x}^{(i)}$. A crucial feature of this dual formulation is that the optimization algorithm only depends on inner products of the training samples $\langle \boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)} \rangle$. In addition, at prediction time, one only needs to calculate the inner product between the training samples and new samples. Because the algorithm is formulated such that the input vector enters only in the form of a scalar product, this allows us to employ the *kernel trick*, whereby we replace $\boldsymbol{x}^{(i)} \cdot \boldsymbol{x}^{(j)}$ with some other choice of kernel function, $K(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)})$. Then, at prediction time the decision function has the form

$$d(\boldsymbol{x}) = \sum_{i=1}^{N} \alpha_i y_i K(\boldsymbol{x}^{(i)}, \boldsymbol{x}) + b, \tag{7.2}$$

whereby we can learn more complex decision functions depending on the choice of the kernel without explicitly generating more features in our input $\boldsymbol{x}$. For example, the kernel $K(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)}) = (\boldsymbol{x}^{(i)} \cdot \boldsymbol{x}^{(j)} + c_0)^d$ corresponds to the mapping to a $\binom{p+d}{d}$ dimensional feature space corresponding to all the monomials of the form $x_{i_1} x_{i_2} \ldots x_{i_k}$ (ignoring permutations) that are up to order d where p is the number of *raw* features. Importantly, we note that the choice of a kernel (e.g. a polynomial) that results in a highly interpretable decision function also limits the expressive power of the SVM, especially when compared with universal approximator algorithms such as neural networks.

7.3 Results

In this section, we perform supervised learning with SVMs on data sets generated by sampling spin configurations of classical Hamiltonians, where Ising degrees of freedom $\boldsymbol{\sigma}$ ("spins" taking binary values) will serve as our input $\boldsymbol{x}$. Finite size lattices with N spins are considered. In the following, we use the most efficient training algorithm for SVMs by means of the scikit-learn library [165] – the Sequential Minimal Optimization algorithm [166] – known to scale as $O(N^2)$, or with a smaller power, for several kernels and types of data. We find, in practice, training on a single core is generally slow for 10^5 samples or more. We explore the behaviour of different polynomial kernels and perform grid search to find the optimal regularization parameter γ. In general, the results are averaged over several choices of training and test sets for the same hyperparameters in order to obtain more reliable statistics.

7.3.1 2d Ising Model

We first consider the nearest-neighbor Ising model in two dimensions, $H = -\sum_{\langle \boldsymbol{ab} \rangle} \sigma_{\boldsymbol{a}} \sigma_{\boldsymbol{b}}$, where $\sigma_{\boldsymbol{a}} = \pm 1$, and $\boldsymbol{a}$ are the Euclidean coordinates of a given lattice site. Monte Carlo simulations using the Wolff algorithm were performed to collect spin configurations $\boldsymbol{\sigma}^{(i)} = (\ldots, \sigma_{\boldsymbol{a}}^{(i)}, \ldots)$ where i identifies each configuration collected at different temperatures from $T = 1.6$ to $T = 2.9$. The 2d Ising phase transition occurs at the critical temperature $T_c \approx 2.269$ [167] and separates a ferromagnetic (FM) phase, characterized by a non-zero total magnetization per spin, from a featureless paramagnetic (PM) phase at high temperatures. For a given $L \times L = N$ size lattice, each sample is labeled with its corresponding phase in the binary class $y_i = \pm 1$. We train SVMs to learn to discriminate between the two phases for different numbers of samples in the training set, with a corresponding change in the number of samples at each temperature. For the Ising model, we limit our survey to a linear and a quadratic kernel of general form $K(\boldsymbol{\sigma}, \boldsymbol{\sigma}') = (\boldsymbol{\sigma} \cdot \boldsymbol{\sigma}' + c_0)^k$ with $k = 1, 2$ and $c_0 = 0$. Note that in general, it might be necessary for the learning procedure to find the optimal c_0 as well.

In order to quantify the performance of each SVM model, the main metric that we study is the test set accuracy as a function of the number of samples in the training set, for the value of γ which results in the optimal accuracy of the model, illustrated in Fig. 7.1 (a). For the linear kernel, the exploration of γ is over a log-spaced grid of 11 values from 10^{-5} to 10^5. For the quadratic kernel the accuracy does not depend significantly on the choice of γ and thus we fixed it at $\gamma = 10^{-5}$. For a given number of training samples and regularization γ, the test set accuracy is additionally averaged over different random selections of training and test sets.

Results for the test set accuracy and for the SVM decision function are shown in Fig. 7.1. As seen in Fig. 7.1 (a), the quadratic kernel performs extremely well with mean test set accuracy $\approx 97\%$ for $L = 40$. This can be easily *interpreted*, since we know that this model possesses a quadratic order parameter that linearly discriminates the FM from the PM, i.e. the squared magnetization per spin $m^2 = (\sum_{\boldsymbol{a}} \sigma_{\boldsymbol{a}}/N)^2$. We find that the quadratic kernel reaches very significant performance with only a few dozen samples in the training set, which is a result of the simplicity of this model. Moreover, with increasing number of samples the test set accuracy approaches a plateau value which increases with system size towards 100%. This is the expected behaviour, since at the critical point the fluctuations of the order parameter approach zero in the thermodynamic limit and it is thus possible to discriminate perfectly between both phases.

For the linear kernel (Fig. 7.1 (a)), the accuracy shows non-monotonic behaviour with

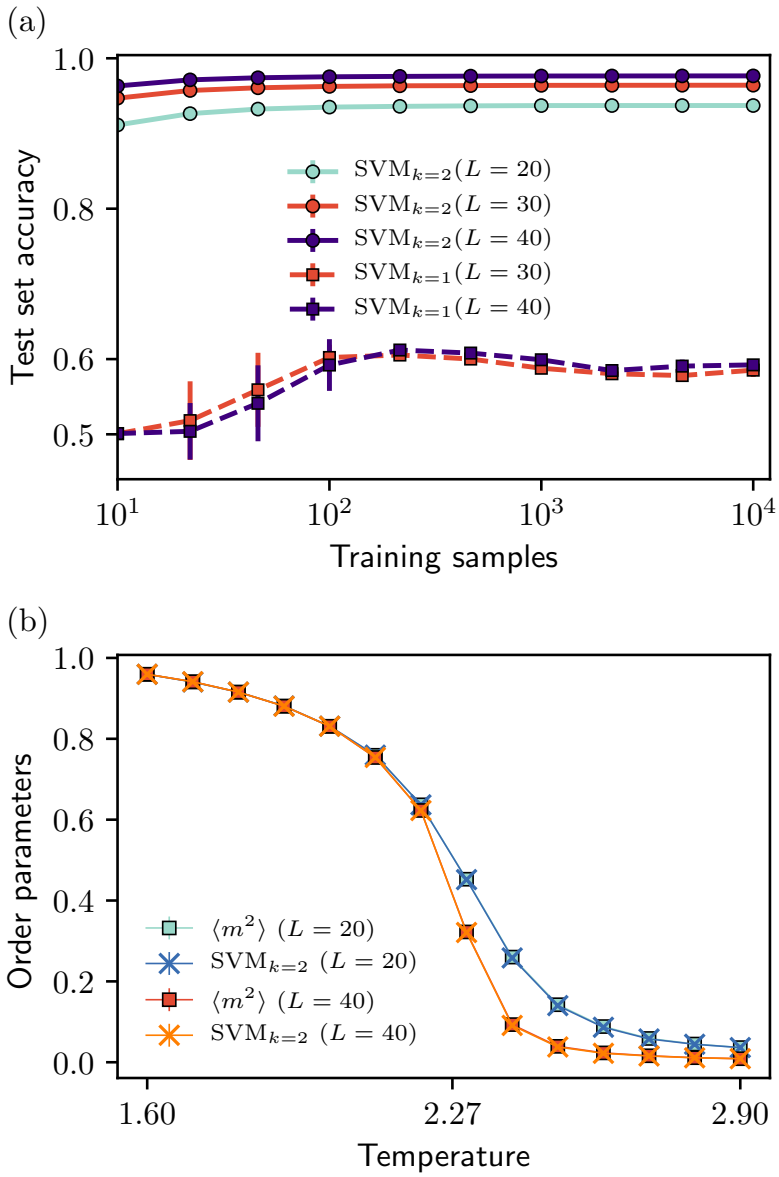

Figure 7.1: (a) Average test set accuracy of Support Vector Machines with polynomial kernel $K(\boldsymbol{\sigma}, \boldsymbol{\sigma}') = (\boldsymbol{\sigma} \cdot \boldsymbol{\sigma}')^k$ trained on Monte Carlo sampled configurations from the 2d Ising model. For each number of training samples, the accuracy is averaged over 100 independent training and test sets. (b) The SVM classifies samples according to $\text{sign}(d(\boldsymbol{\sigma}))$. The decision function $d(\boldsymbol{\sigma})$ for the SVM with a quadratic polynomial kernel is evaluated by Monte Carlo sampling at different temperatures and compared to the squared magnetization per spin m^2. The arbitrary scale factor and off-set in the SVM decision function are fixed by matching the decision function to $\langle m^2 \rangle$ at $T = 1.6$ and $T = 2.9$.

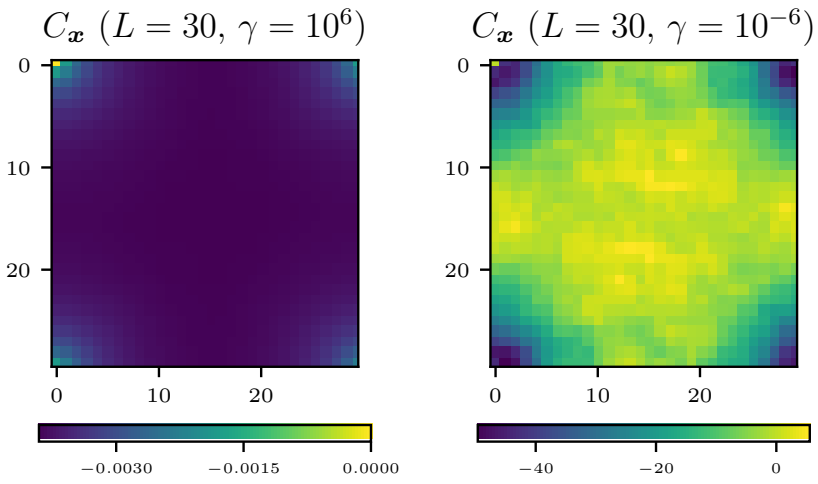

Figure 7.2: The decision function for an SVM with quadratic polynomial kernel, Eq. 7.3. The averaged $C_{\boldsymbol{x}} = \langle C_{\boldsymbol{x}}^{(\boldsymbol{a})} \rangle_{\boldsymbol{a}}$ is displayed for different values of regularization $\gamma = 10^6$ and $\gamma = 10^{-6}$. Clearly, for large regularization ($\gamma = 10^6$), the decision function is essentially the m^2 order parameter of the 2d Ising model.

the total number of training set samples and does not improve with increasing system size. This is a consequence of the fact that a linear decision function is unable to discriminate between the FM and PM phases. Namely, in the FM phase configurations have magnetization per spin near ± 1, while for the PM phase most configurations have appproximately zero magnetization. Thus, the $k = 1$ kernel is asking a linear decision boundary to separate a data set with three clusters – an impossible task. Close inspection of the decision function learned by the SVM reveals it contains random linear coefficients without any structure, confirming that nothing physically relevant is being learned about the data in this case.

As noted above, the accuracy of the SVM with a quadratic kernel on the test set does not depend significantly on the regularization parameter γ. An advantage of SVMs is that we can visualize the decision function being learned. From Eq. 7.2, the decision function for an SVM with quadratic polynomial kernel can be expressed as

$$d(\boldsymbol{\sigma}) = \sum_{\boldsymbol{a}} \sum_{\boldsymbol{x}} C_{\boldsymbol{x}}^{(\boldsymbol{a})} \boldsymbol{\sigma}_{\boldsymbol{a}} \boldsymbol{\sigma}_{\boldsymbol{a}+\boldsymbol{x}} + b. \tag{7.3}$$

In Fig. 7.2, we display the heatmap of $C_{\boldsymbol{x}} = \langle C_{\boldsymbol{x}}^{(\boldsymbol{a})} \rangle_{\boldsymbol{a}}$, where $\langle \ldots \rangle_{\boldsymbol{a}}$ denotes averaging with respect to all sites $\boldsymbol{a}$ for $\gamma = 10^6$ and $\gamma = 10^{-6}$ and system size $L = 30$. It is

interesting to note that even though the classification performance is very similar, the SVM decision function corresponds to different order parameters depending on the amount of regularization. Clearly, at $\gamma = 10^6$, the SVM is learning m^2 as the order parameter of the model up to finite-size effects. In contrast, at $\gamma = 10^{-6}$, the SVM is learning to calculate the square of the total magnetization within some fixed distance of each spin and summing all these different local contributions. A more detailed Fourier decomposition of $C_{\boldsymbol{x}}^{(\boldsymbol{a})}$ shows that the $\boldsymbol{k} = 0$ mode is dominant for any amount of regularization, but at small γ the contributions from other small $\boldsymbol{k}$ modes are larger.

In Fig. 7.1 (b), the SVM decision function (for large regularization $\gamma = 10^6$) is averaged over Monte Carlo samples at different temperatures showing essentially perfect agreement with m^2. Of course, the SVM decision function has an arbitrary scale and off-set and in order to match with $\langle m^2 \rangle$ a linear transformation is performed, so that they agree at the extreme values of temperature $T = 1.6$ and 2.9.

7.3.2 Conserved-order-parameter Ising model

The conserved-order-parameter Ising model [135] is described by the same Hamiltonian as the Ising model but the configuration space is restricted to the subspace where the total magnetization $\sum_{\boldsymbol{a}} \sigma_{\boldsymbol{a}}$ is zero. This model describes a half-filled lattice gas of particles with a nearest neighbour attractive interaction [168]. At low temperatures, domains of up and down spins are separated by either two horizontal or two vertical domain walls for a square lattice geometry with periodic boundary conditions. There is a phase transition to a featureless phase at the same critical temperature as the 2d Ising model. Ref. [135] studied this model on the square lattice using Principal Component Analysis (PCA), a dimensional reduction algorithm, and devised the following order parameter:

$$\begin{aligned} S = & \frac{1}{L^4} \sum_{\boldsymbol{a},\boldsymbol{b}} \sigma_{\boldsymbol{a}} \sigma_{\boldsymbol{b}} \left[\cos\left(\frac{2\pi}{L} (a_1 - b_1) \right) \right. \\ & \left. + \cos\left(\frac{2\pi}{L} (a_2 - b_2) \right) \right]. \end{aligned} \tag{7.4}$$

S is of the form of Eq. 7.3 with

$$C_{\boldsymbol{x}}^{(\boldsymbol{a})} = \frac{1}{L^4} \left[\cos\left(\frac{2\pi}{L} x_1 \right) + \cos\left(\frac{2\pi}{L} x_2 \right) \right]. \tag{7.5}$$

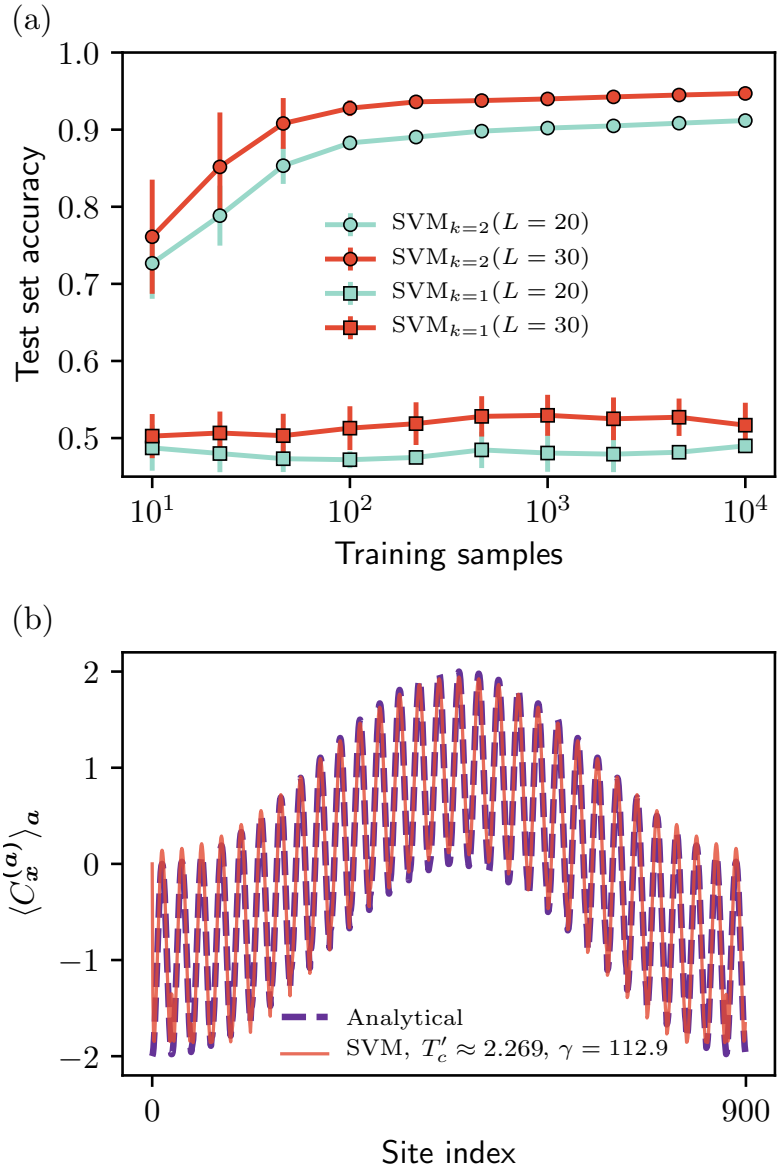

Figure 7.3: (a) Average test set accuracy vs number of training samples for Support Vector Machines with polynomial kernel $K(\boldsymbol{\sigma}, \boldsymbol{\sigma}') = (\boldsymbol{\sigma} \cdot \boldsymbol{\sigma}')^k$ trained on Monte Carlo sampled configurations from the conserved-order-parameter Ising model at different temperatures. (b) The spatial dependence of the SVM decision function coefficients $\langle C_{\boldsymbol{x}}^{(\boldsymbol{a})} \rangle_{\boldsymbol{a}}$ learned by an SVM with quadratic kernel and regularization coefficient $\gamma = 112.9$ for system size $L = 30$ shows very good agreement with the analytical form (Eq. 7.5) devised in [135].

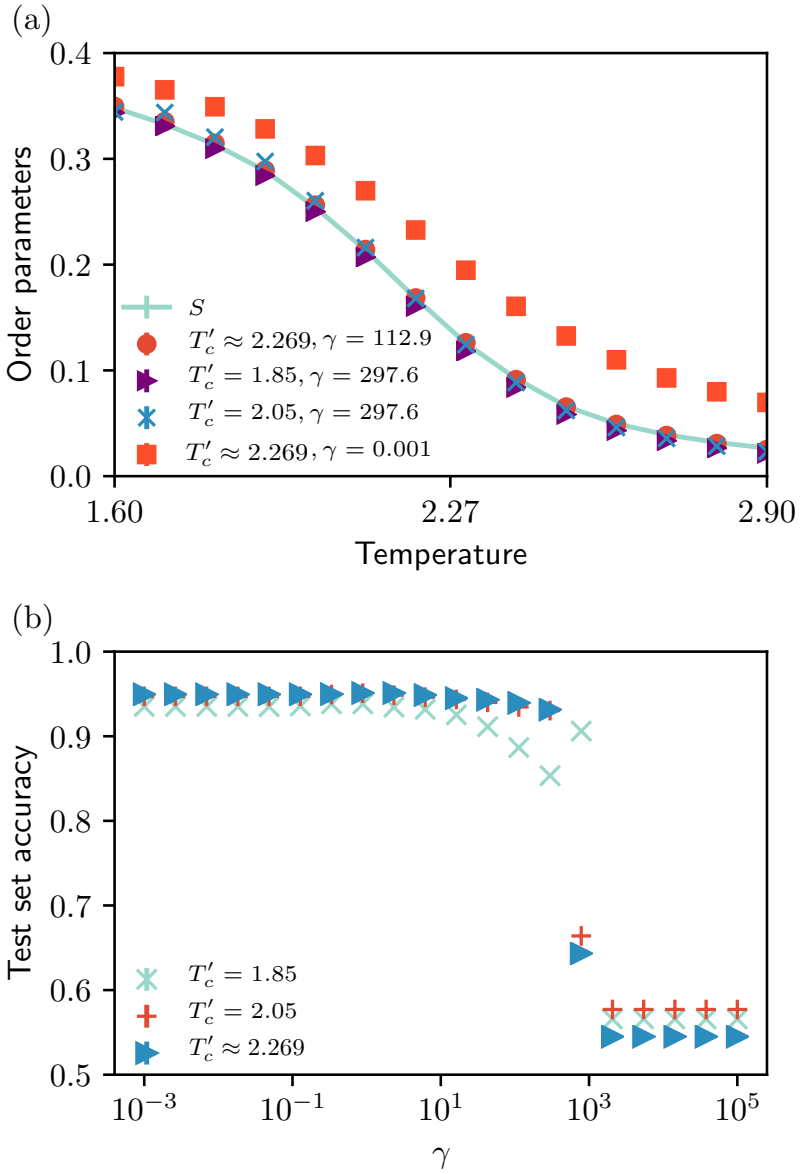

Figure 7.4: (a) Comparing the decision function averaged over Monte Carlo samples at different temperatures learned by SVMs with quadratic kernel - trained on 30000 samples with $L = 30$ - to the analytical order parameter S (Eq. 7.4) assuming different critical temperatures T_c' and amount of regularization γ. (b) Dependence on regularization γ of the test set accuracy (averaged over 10 sets of 10000 samples) of quadratic SVMs trained on 30000 samples for $L = 30$.

In contrast to PCA, which through feature dimension reduction provides the easy visualization required for the determination of this order parameter, SVMs should provide automated order parameter detection in a more systematic way. We explore this ability now.

As in the case with the Ising model, we collect Monte Carlo samples for a set of temperatures below and above T_c. Fig. 7.3 (a) shows the averaged test set accuracy as a function of the number of samples for the optimal value of the regularization parameter γ over a grid of 11 values from 10^{-5} to 10^5. The results are averaged over 400 training and test sets. As in the Ising model case, for a quadratic kernel there is a monotonic improvement of test accuracy with the number of training samples. The limiting value as the number of training samples becomes very large also increases with system size. This behavior signals the existence of a quadratic order parameter that discriminates between the two phases. As a check, we compare the explicit spatial dependence $C_{\boldsymbol{x}} = \langle C_{\boldsymbol{x}}^{(\boldsymbol{a})} \rangle_{\boldsymbol{a}}$ of the SVM decision function for $\gamma = 112.9$ (the selection of this regularization value is discussed below) to the analytical form (Eq. 7.5). The scale and off-set of the SVM decision function are fixed by requiring the coefficients $C_{\boldsymbol{x}}$ to have the same mean and standard deviation as the analytical form. Fig. 7.3 (b) shows very good agreement between them. In contrast, the linear kernel does not show a clear improvement of the test set accuracy with increasing number of samples which indicates there isn't an order parameter of that form.

One could ask whether this order parameter could also be learned in the case where the precise value of the critical temperature T_c' is not known. We address this by performing supervised learning with quadratic SVMs assuming different values T_c' for the critical temperature. Finite size scaling of the learned order parameter (or its moments) can then be used to estimate the value of the critical temperature. Fig. 7.4 (a) compares the actual order parameter S (Eq. 7.4) with the SVM decision function for different values of $T_c' = 1.85$, 2.05, and $T_c' \approx 2.269$. The scale and off-set of the SVM decision function are fixed by matching it with S at $T = 0.1$ and $T = 100.0$. We observe that the SVM decision functions learned when assuming $T_c' = 1.85, 2.05$ (for a choice of regularization values discussed below) also agree well with S, suggesting machine learning algorithms are able to learn important physical information without the precise knowledge of T_c.

We now analyze the role of the regularization parameter γ for learning in this model. Fig. 7.4 (b), shows for $L = 30$ and a training set with 30000 samples, the test set accuracy averaged over 10 randomly picked test sets of 10000 samples versus the regularization parameter γ for $T_c' = 1.85$, 2.05, and $T_c' \approx 2.269$. At $\gamma \lesssim 10^3$, the test set accuracy jumps significantly to very high values ($\approx 90\%$) and then reaches a plateau. Interestingly, we find that throughout the plateau region the decision function of the SVM changes continuously.

Fig. 7.4 (a) also compares the analytical order parameter S with the SVM decision function for different values of γ when $T_c' = T_c$. For $\gamma = 112.9$, right at the beginning of the plateau, there is highly accurate agreement between S and the SVM decision function, while for $\gamma = 0.001$ the two do not match. Similar results were also found for the other values of T_c'. Thus, the physical order parameter can be interpreted as being associated with the least amount of complexity that still allows the model to have good predictive performance. This was also what we found for the Ising model in Section 7.3.1.

7.3.3 2d Ising Gauge theory

Finally, we consider the challenging case of topologically ordered systems, where no conventional local parameter exists. For concreteness we study the 2d Ising gauge theory with Hamiltonian $H = -\sum_p \prod_{i \in p} \sigma_i^z$, where the spins live on the bonds of a 2d square lattice and p represents a plaquette with four spins. The set of ground states is a degenerate manifold with the constraint that for all plaquettes p, $\prod_{i \in p} \sigma_i^z = 1$. In the thermodynamic limit, the constraints are violated at any finite temperature. A conventional order parameter that distinguishes ground states from finite temperature states does not exist. Ref. [134] found that the simplest fully connected feed forward neural networks were unable to classify these two different cases in a supervised learning context. Only with convolutional neural networks, which explicitly exploit locality and translational invariance, did this classification task succeed on raw spin configurations. We note that by engineering predictive features as a pre-processing step, simple feed forward neural networks may be successful in classification (as demonstrated in Ref. [156]). However in the present case, we explore the behaviour of SVMs for the raw 2d Ising Gauge theory data with no pre-processing on the input data.

We perform supervised learning on spin configurations generated for the Ising Gauge theory at $T = 0$ and $T = \infty$ (i.e. completely random spin states). We explore kernels of increasing degree, starting from $k = 1$, for different system sizes L. Fig. 7.5 (a) shows that for system size $L = 5$, SVMs with polynomial kernel of degree less than 4 fail to discriminate these two phases, exhibiting average test set accuracies $\sim 50\%$, which amounts to random guessing. However, a SVM with polynomial kernel of fourth order is able to perform the task with an accuracy that converges towards 100% as the number of training samples is increased. As apparent in Fig. 7.5 (b), we also note that the number of training samples necessary to learn the fourth order discriminator increases with system size. This illustrates the difficulty that SVMs can have in performing classification for large system sizes.

In order to interpret these results, we further analyze the decision function of the SVM.

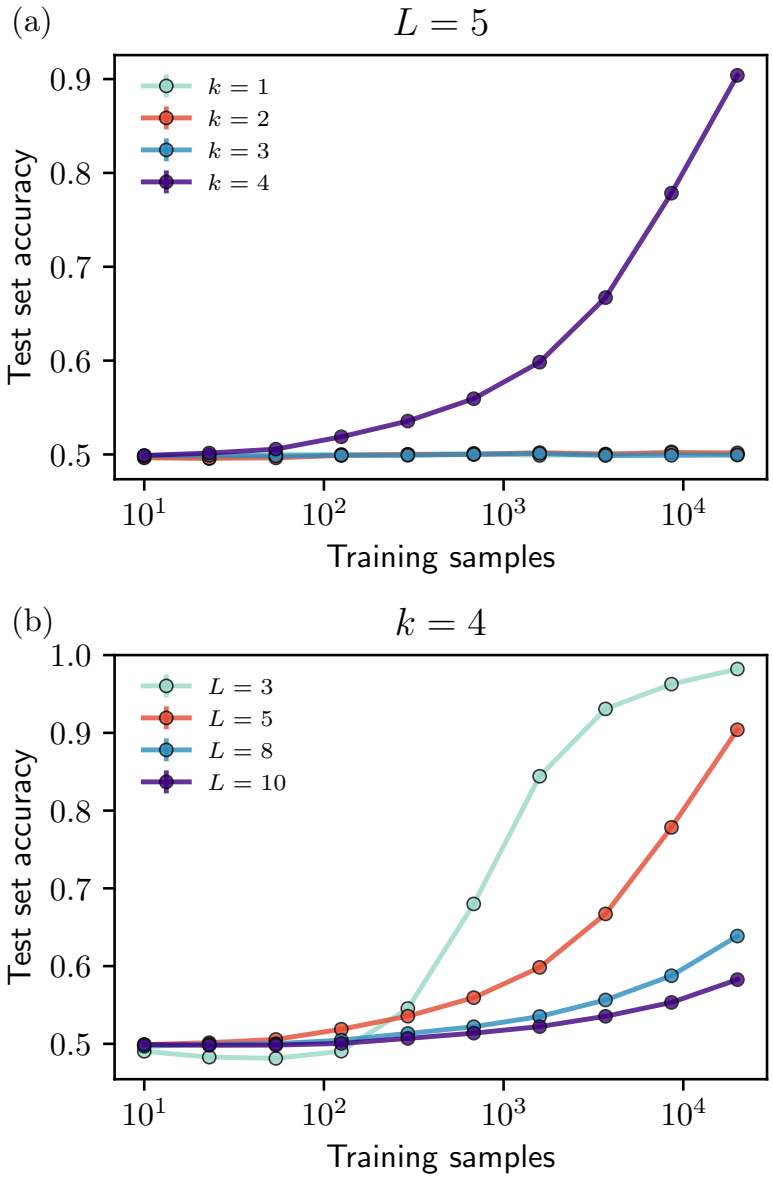

Figure 7.5: (a) Test set accuracy of SVMs with polynomial kernels of order k in classifying ground states versus infinite temperature states of the 2d Ising Gauge theory for system size $L = 5$. Only in the feature space of fourth order polynomials is the SVM algorithm performance better than a random algorithm. The performance approaches 100% with increasing number of training samples. (b) Test set accuracy for $k = 4$ and different system sizes. For larger system sizes, more training samples are needed to learn the correct decision function.

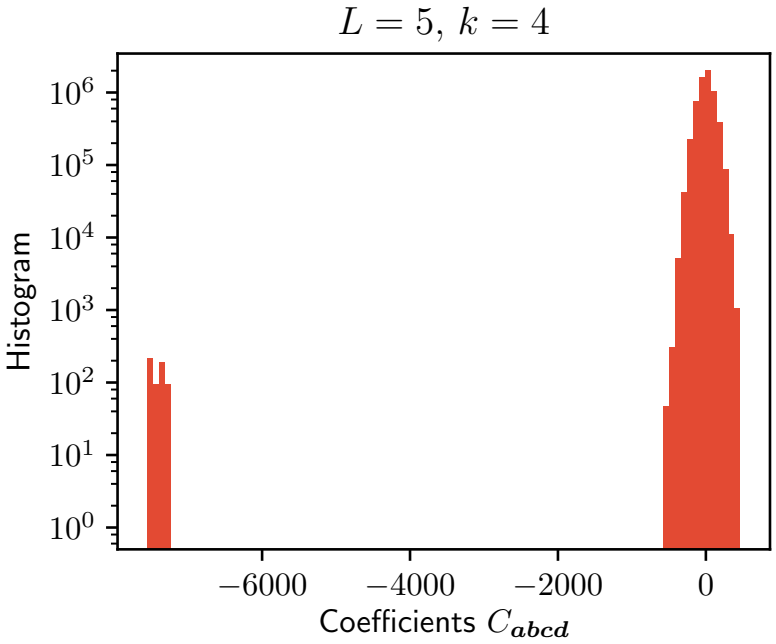

Figure 7.6: Histogram of the coefficients $C_{\boldsymbol{abcd}}$ in the decision function learned by a SVM with 4th order polynomial kernel in classifying ground states versus infinite temperature states of the 2d Ising Gauge theory for $L = 5$. There are 600 large coefficients (in absolute value) which correspond to the 25 plaquettes and the permutations of their spins.

The smallest degree (fourth order) polymonial kernel learns a decision function of the form $d(\boldsymbol{\sigma}) = \sum_{\boldsymbol{abcd}} C_{\boldsymbol{abcd}} \sigma_{\boldsymbol{a}} \sigma_{\boldsymbol{b}} \sigma_{\boldsymbol{c}} \sigma_{\boldsymbol{d}} + b$, which contains the product of four spins. Fig. 7.6 shows the histogram of the coefficients $C_{\boldsymbol{abcd}}$ for an SVM trained on 20000 samples and $\gamma = 10^{-6}$. From this plot it is clear that there are two sets of coefficients; one near zero, and another set with large negative values. Counting the number of these large (in magnitude) coefficients reveals 600. For this lattice, with 25 square plaquettes, this number corresponds to the the number of possible permutations of the four indices of $\boldsymbol{abcd}$, i.e. $25 \times 4!$. Hence, this model provides evidence that SVMs are able to learn complex interpretable decision functions and discover the locality of the Hamiltonian directly from raw data on spin configurations.

7.4 Conclusions

We have examined the use of support vector machines (SVMs), one of the most common tools for supervised learning, for the binary classification of phases in several models of interest to condensed matter and statistical physics. SVMs employ a *kernel trick* to define a decision function, used to discriminate features in a higher dimensional space. The kernel depends on the inner product of spin vectors, but can otherwise have some freedom in definition. We have shown that in contrast to other methods such as neural networks, the ability to use different kernels for classification tasks gives SVMs significant value in finding *interpretable* physical discriminators for different models, such as conventional order parameters defined in condensed matter theory.

To allow for some slack to misclassify data, SVMs employ a regularization, which controls the tradeoff between minimizing training errors, while still allowing sufficient model complexity. In this chapter we have found that the amount of regularization has a strong impact on the decision function learned. We find that the expected physical order parameter is associated with the least complex model (i.e. the largest amount of regularization) that is still able to obtain near optimal test set accuracy.

On the other hand, we find that small values of regularization that achieve the same (or better) performance than the physical order parameter learn a non-physical decision function – one not related to any conventional order parameter. Since such decision functions generalize well to uncorrelated test set samples, we argue that this is not an example of overfitting, even though it may arise due to particular details contained within the training set. This observation deserves further study, as it may have consequences more generally for the role of regularization in black box algorithms, such as neural networks, when applied to data in condensed matter physics.

While very successful on the Ising-like examples studied in this chapter, the SVM algorithm is not without its limitations. As observed for the 2d Ising Gauge Theory, the number of samples required to learn the physically relevant decision function can grow prohibitively large. The reason that this occurs for the degenerate groundstate of the Ising gauge theory, and not the models with conventional order parameters, desires further study. We generally observe that the training of SVMs when the number of samples is larger than 10^5 becomes time consuming. This could also be an issue in using SVMs near phase transitions in models where a large number of samples is necessary to learn the physical discriminator. In such cases, it would be interesting to further explore the interpretability of neural networks, since they can possibly present better scalability to larger system sizes.

It would be interesting to study the utility of SVMs on other classical models of interest

in condensed matter physics. We note that for systems with continuous degrees of freedom it might be necessary to consider a radial basis function as a kernel, which maps to an infinite dimensional space, instead of polynomials. For example, it would interesting to see if for the 2d XY model, the spin stiffness can be identified as the physical discriminator for the Kosterlitz-Thouless transition using a suitably modified SVM. Finally, SVMs and other sparse kernel machines may in the future be easily adapted to study quantum phases and phase transitions by using wavefunctions or density matrices as data. SVMs with linear or quadratic polynomial kernels could be used to determine operators which discriminate between different quantum phases using density matrices or wavefunctions as data. Thus, as the condensed matter community increasingly adopts modern machine learning methods into its numerical arsenal, we expect that SVMs will become a standard tool for finding *interpretable* physical discriminators for generic Hamiltonians in the near future.

Chapter 8

Discussion and Outlook

In the previous chapter, we demonstrated that it is possible to automate through machine learning the discovery of important interpretable features associated with phase transitions such as order parameters.

In the context of this thesis, a natural future research direction concerns the application of interpretable machine learning to phase transitions such as the MBL-to-thermal transition that are not well understood theoretically. The results of Ref. [143] provide hopeful prospects as it showed that the decision function associated with a trained neural network was able to discriminate MBL phases with different types of order and also the MBL-to-thermal transition with sharper crossovers than standard measures such as the standard deviation of bipartite entanglement entropy. This suggests the existence of so far unknown physical quantities that might be relevant for a better understanding of this phase transition. Another interesting research direction in this context regards a data-driven identification of the thermal inclusions which are believed to play an important role at the transition. This could possibly be achieved through unsupervised machine learning algorithms (e.g. clustering).

The field of learning phase transitions with machine learning is currently not without its challenges. For example, supervised learning with neural networks of different architectures cannot capture the topological features associated with the Kosterlitz-Thouless transition in the 2d XY model without explicit feature engineering [169]. Instead, it seems to learn the magnetization (which is non-zero due to finite size effects at the transition temperature for accessible system sizes) and energy fluctuations as a way to detect the phase transition. It is an interesting open question whether the cost function of the neural network could be modified to bias it towards finding the true order parameter.

Challenges such as these are important milestones for the development of the field. In the future, as the field matures through the application of machine learning to various models including those that are well understood from a theory point of view and those that are not, machine learning algorithms might become standard tools not only to discover phase transitions, but also to guide their theoretical understanding.

Bibliography

[1] P. Ponte, A. Chandran, Z. Papic, and D. A. Abanin, Ann. Phys. (N. Y). 353, 196 (2015).

[2] P. Ponte, Z. Papic, F. Huveneers, and D. A. Abanin, Phys. Rev. Lett. 114, 140401 (2015).

[3] P. Ponte, C. R. Laumann, D. A. Huse, and A. Chandran, Phil. Trans. R. Soc. A 375, 20160428 (2017).

[4] P. Ponte and R. G. Melko, Phys. Rev. B 96, 205146 (2017).

[5] L.D. Landau and E.M. Lifshitz, Statistical Physics, Third Edition, Part 1, Volume 5, Butterworth-Heinemann, 1980.

[6] L. D'Alessio, Y. Kafri, A. Polkovnikov, and M. Rigol, Adv. in Phys. 65, 239 (2016).

[7] M. Rigol, V. Dunjko, and M. Olshanii, Nature 452, 854 (2008).

[8] A. Polkovnikov, K. Sengupta, A. Silva, and M. Vengalattore, Rev. Mod. Phys. 83, 863 (2011).

[9] J. M. Deutsch, Phys. Rev. A 43, 2046 (1991).

[10] M. Srednicki, Phys. Rev. E 50, 888 (1994).

[11] M. Srednicki, Journal of Physics A: Mathematical and General 32, 1163 (1999).

[12] E. Khatami, G. Pupillo, M. Srednicki, and M. Rigol, Phys. Rev. Lett. 111, 050403 (2013).

[13] R. Steinigeweg, J. Herbrych, and P. Prelovsek, Phys. Rev. E 87, 012118 (2013).

[14] W. Beugeling, R. Moessner, and M. Haque, Phys. Rev. E 91, 012144 (2015).

[15] H. Kim, T. N. Ikeda, and D. A. Huse, Phys. Rev. E 90, 052105 (2014).

[16] A. Chandran, V. Khemani, C. R. Laumann, and S. L. Sondhi, Phys. Rev. B 89, 144201 (2014).

[17] S. Sachdev, Quantum Phase Transitions (Cambridge University Press, 2nd ed., 2007).

[18] M. Rigol, V. Dunjko, V. Yurovsky, and M. Olshanii, Phys. Rev. Lett. 98, 050405 (2007).

[19] S. Ziraldo and G. E. Santoro, Phys. Rev. B 87, 064201 (2013).

[20] L. Vidmar and M. Rigol, J. Stat. Mech. 064007 (2016).

[21] R. Nandkishore and D. A. Huse, Annu. Rev. Condens. Matter Phys. 6, 15 (2015).

[22] D. A. Abanin and Z. Papic, Annalen der Physik 529, 1700169 (2017).

[23] D. M. Basko, I. L. Aleiner, and B. L. Altshuler, Ann. Phys. 321, 1126 (2006).

[24] I. V. Gornyi, A. D. Mirlin, and D. G. Polyakov, Phys. Rev. Lett. 95, 206603 (2005).

[25] M. Serbyn, Z. Papic, and D. A. Abanin, Phys. Rev. Lett. 111, 127201 (2013).

[26] D. A. Huse, R. Nandkishore, and V. Oganesyan, Phys. Rev. B 90, 174202 (2014).

[27] J. H. Bardarson, F. Pollmann, and J. E. Moore, Phys. Rev. Lett. 109, 017202 (2012).

[28] H. Kim, D. A. Huse, Phys. Rev. Lett. 111, 127205 (2013).

[29] P. W. Anderson, Phys. Rev. 109, 1492 (1958).

[30] A. Pal and D. A. Huse, Phys. Rev. B 82, 174411 (2010).

[31] D. J. Luitz, N. Laflorencie, and F. Alet, Phys. Rev. B 91, 081103 (2015).

[32] V. Oganesyan and D. A. Huse, Phys. Rev. B 75, 155111 (2007).

[33] M. L. Mehta, Random Matrix Theory, Springer, New York (1990).

[34] F. Haake, Quantum Signatures of Chaos, Vol. 54 (Springer, 2010).

[35] Y. Y. Atas, E. Bogomolny, O. Giraud, and G. Roux, Phys. Rev. Lett. 110, 084101 (2013).

[36] Michael A. Nielsen and Isaac L. Chuang, Quantum Computation and Quantum Information (Cambridge University Press, 2011).

[37] W. Beugeling, A. Andreanov, and M. Haque, J. Stat. Mech. 2015, P02002 (2015).

[38] D. Delande, Kicked rotor and Anderson localization, Boulder School on Condensed Matter Physics (2013).

[39] G. Casati and J. Ford, *Stochastic Behavior in Classical and Quantum Hamiltonian Systems*, 1st ed., Vol. **93** (Springer Berlin Heidelberg, 1979).

[40] D. R. Grempel, R. E. Prange, and S. Fishman, Phys. Rev. A 29, 1639 (1984).

[41] F. L. Moore, J. C. Robinson, C. Bharucha, P. E. Williams, and M. G. Raizen, Phys. Rev. Lett. 73, 2974 (1994).

[42] G. Lemarié, J. Chabé, P. Szriftgiser, J. C. Garreau, B. Grémaud, and D. Delande, Phys. Rev. A 80, 043626 (2009).

[43] T. Dittrich and R. Graham, Annals Phys. 200, 363 (1990).

[44] F. M. Izrailev and D. L. Shepelyanskii, Theor. Math. Phys. 43, 1586 (1980).

[45] Fritz Haake, Quantum signatures of chaos, Springer-Verlag, Berlin, 2nd edition (1991).

[46] M. Bukov, L. DAlessio, and A. Polkovnikov, Advances in Physics 64, 139 (2015).

[47] I. Bloch, J. Dalibard, and W. Zwerger, Rev. Mod. Phys. 80, 885 (2008).

[48] B. P. Abbott et al. (Virgo, LIGO Scientific), Phys. Rev. Lett. 116, 061102 (2016).

[49] M. Levitt, Spin Dynamics: Basics of Nuclear Magnetic Resonance, John Wiley & Sons (2001).

[50] T. Kitagawa, T. Oka, A. Brataas, L. Fu, and E. Demler, Phys. Rev. B 84, 235108 (2011).

[51] A. Lazarides, A. Das, and R. Moessner, Phys. Rev. Lett. 112, 150401 (2014).

[52] A. Polkovnikov, Quantum ergodicity: fundamentals and applications, http://physics.bu.edu/~asp/teaching/PY_747.pdf (2013).

[53] W. Magnus, Pure Appl. Math 7 (1954) 649.

[54] S. Blanes, F. Casas, J. A. Oteo, and J. Ros, Phys. Rep. 470, 151 (2009).

[55] J. H. Shirley, Phys. Rev. 138, B979 (1965).

[56] L. D'Alessio, and A. Polkovnikov, Annals of Physics 333 (2013).

[57] D. R. Grempel, R. E. Prange, and S. Fishman, Phys. Rev. A 29 (1984).

[58] L. DAlessio and M. Rigol, Phys. Rev. X 4, 041048 (2014).

[59] N. Regnault and R. Nandkishore, Phys. Rev. B 93, 104203 (2016).

[60] A. Lazarides, A. Das, and R. Moessner, Phys. Rev. Lett. 112, 150401 (2014).

[61] A. Lazarides, A. Das, and R. Moessner, Phys. Rev. E 90, 012110 (2014).

[62] I. Bloch, J. Dalibard, and W. Zwerger, Rev. Mod. Phys. 80 (2008).

[63] R. Blatt, and C. F. Roos, Nature Physics 8 277 (2012).

[64] P. Neumann, N. Mizuochi, F. Rempp, P. Hemmer, H. Watanabe, S. Yamasami, V. Jacques, T. Gaebel, F. Jelezko, and J. Wrachtrup, Science 320 1326 (2008).

[65] S. S. Kondov, W. R. McGehee, J. J. Zirbel, and B. DeMarco, Science 334 (6052) (2011).

[66] M. Pasienski, D. McKay, M. White, and B. DeMarco, Nat. Phys. 6, 677 (2010).

[67] B. Deissler, M. Zaccanti, G. Roati, C. DErrico, M. Fattori, M. Modugno, G. Modugno, and M. Inguscio, Nat. Phys. 6, 354 (2010).

[68] S. S. Kondov, W. R. McGehee, W. Xu, and B. DeMarco, Phys. Rev. Lett. 114, 083002 (2015).

[69] A. Chandran, I. H. Kim, G. Vidal, and D. A. Abanin, Phys. Rev. B 91, 085425 (2015).

[70] V. Ros, M. Mller, and A. Scardicchio, Nucl. Phys. B891, 420 (2015).

[71] B. Bauer and C. Nayak, J. Stat. Mech. P09005 (2013)

[72] D. N. Page, Phys. Rev. Lett. 71, 1291 (1993).

[73] J. A. Kjäll, J. H. Bardarson, and F. Pollmann, Phys. Rev. Lett. 113, 107204 (2014).

[74] R. B. Sidje, ACM Trans. Math. Softw., **24**, 130 (1998).

[75] G. Vidal, Phys. Rev. Lett. 91, 147902 (2003).

[76] R. Vasseur, S. A. Parameswaran, and J. E. Moore, Phys. Rev. B 91, 140202 (2015).

[77] M. Serbyn, Z. Papic, and D. A. Abanin, Phys. Rev. B 90, 174302 (2014).

[78] M. Serbyn, Z. Papić, and D. A. Abanin, Phys. Rev. Lett. 110, 260601 (2013).

[79] M. Znidaric, T. Prosen, and P. Prelovsek, Phys. Rev. B 77, 064426 (2008).

[80] P. Calabrese and J. Cardy, J. Stat. Mech. P04010 (2005).

[81] G. De Chiara, S. Montangero, P. Calabrese, and R. Fazio, J. Stat. Mech. P03001 (2006).

[82] E. H. Lieb and D. Robinson, Commun. Math. Phys. 28, 251 (1972).

[83] J. Z. Imbrie, Journal of Statistical Physics 163, 998 (2016).

[84] A. Chandran, A. Pal, C. R. Laumann, and A. Scardicchio, Phys. Rev. B 94, 144203 (2016).

[85] W. De Roeck and F. Huveneers, Phys. Rev. B 95 (2017).

[86] X. Yu, D. J. Luitz, and B. K. Clark, Phys. Rev. B 94, 184202 (2016).

[87] V. Khemani, S. P. Lim, D. N. Sheng, and D. A. Huse, Phys. Rev. X 7, 021013 (2017).

[88] D. J. Luitz, F. Huveneers, and W. De Roeck, Phys. Rev. Lett. 119, 150602 (2017).

[89] C. R. Laumann, A. Pal, and A. Scardicchio, Phys. Rev. Lett. 113, 200405 (2014).

[90] A. D. Luca and A. Scardicchio, Europhys. Lett. 101, 37003 (2013).

[91] B. L. Altshuler, Y. Gefen, A. Kamenev, and L. S. Levitov, Phys. Rev. Lett. 78, 2803 (1997).

[92] J. T. Chayes, L. Chayes, D. S. Fisher, and T. Spencer, Phys. Rev. Lett. 57, 2999 (1986).

[93] A. Chandran, C. R. Laumann, and V. Oganesyan, arXiv:1509.04285 (2015).

[94] R. Vosk, D. A. Huse, and E. Altman, Phys. Rev. X 5, 031032 (2015).

[95] A. C. Potter, R. Vasseur, and S. A. Parameswaran, Phys. Rev. X 5, 031033 (2015).

[96] C. Borgs, J. T. Chayes, H. Kesten, and J. Spencer, Communications in Mathematical Physics 224, 153 (2001).

[97] Markus Heydenreich and Remco van der Hofstad. Progress in high-dimensional percolation and random graphs. *Lecture notes for the CRM-PIMS Summer School in Probability*, (2015).

[98] Béla Bollobás. *Random Graphs*. Cambridge University Press, Cambridge, UK, (2001).

[99] R. Abou-Chacra, D. J. Thouless, and P. W. Anderson, Journal of Physics C: Solid State Physics 6, 1734 (1973).

[100] A.L. Burin, arXiv:1610.00811 (2016).

[101] C. Baldwin, C. Laumann, A. Pal, and A. Scardicchio, Phys. Rev. Lett. 118, 127201 (2017).

[102] G. Biroli, A. Ribeiro-Teixeira, and M. Tarzia, arXiv:1211.7334 (2012).

[103] A. De Luca, B. L. Altshuler, V. E. Kravtsov, and A. Scardicchio, Phys. Rev. Lett. 113, 046806 (2014).

[104] B. L. Altshuler, E. Cuevas, L. B. Ioffe, and V. E. Kravtsov, Phys. Rev. Lett. 117, 156601 (2016).

[105] D. S. Fisher, Phys. Rev. B 51, 6411 (1995).

[106] F. Pazmandi, R. T. Scalettar, and G. T. Zimanyi, Phys. Rev. Lett. 79, 5130 (1997).

[107] https://en.wikipedia.org/wiki/Bethe_lattice.

[108] D. A. Abanin, W. D. Roeck, and Francois Huveneers, Ann. Phys. 372, 1 (2016).

[109] P. Bordia, H. Lüschen, U. Schneider, M. Knap, and I. Bloch, Nat. Phys. 13, 460 (2017).

[110] D. A. Huse, R. Nandkishore, V. Oganesyan, A. Pal, and S. L. Sondhi, Phys. Rev. B 88, 014206 (2013).

[111] V. Khemani, A. Lazarides, R. Moessner, and S. L. Sondhi, Physical Review Letters 116, 250401 (2016).

[112] D. V. Else, B. Bauer, and C. Nayak, Phys. Rev. Lett. 117, 090402 (2016).

[113] J. Zhang, P. W. Hess, A. Kyprianidis, P. Becker, A. Lee, J. Smith, G. Pagano, I. D. Potirniche, A. C. Potter, A. Vishwanath, N. Y. Yao, and C. Monroe, Nature 543, 217 (2017).

[114] S. Choi, J. Choi, R. Landig, G. Kucsko, H. Zhou, J. Isoya, F. Jelezko, S. Onoda, H. Sumiya, V. Khemani, C. von Keyserlingk, N. Y. Yao, E. A. Demler, and M. D. Lukin, Nature 543, 221 (2017).

[115] D. A. Abanin, W. De Roeck, and F. Huveneers, Phys. Rev. Lett. 115, 256803 (2015).

[116] D. A. Abanin, W. De Roeck, W. W. Ho, and F. Huveneers, Communications in Mathematical Physics, 1 (2015).

[117] D. A. Abanin, W. De Roeck, W. W. Ho, and F. Huveneers, Phys. Rev. B 95, 014112 (2017).

[118] F. Machado, G. D. Meyer, D. V. Else, C. Nayak, and N. Y. Yao, arXiv:1708.01620 (2017).

[119] T. Mori, T.N. Ikeda, E. Kaminishi, and M. Ueda, arXiv:1712.08790 (2017).

[120] T. Thiery, M. Müller, and W. De Roeck, arXiv:1711.09880 (2017).

[121] N. Bostrom, Superintelligence: The Coming Machine Intelligence Revolution, Oxford UP, 2013.

[122] G. Hinton, et al., Deep neural networks for acoustic modeling in speech recognition. IEEE Signal Processing Magazine 29, 8297 (2012).

[123] A. Krizhevsky, I. Sutskever, G. Hinton, ImageNet classification with deep convolutional neural networks. In Proc. Advances in Neural Information Processing Systems 25 10901098 (2012).

[124] Litjens, Geert, et al., A survey on deep learning in medical image analysis, arXiv:1702.05747 (2017).

[125] Kaggle, Higgs boson machine learning challenge, https://www.kaggle.com/c/higgs-boson (2014)

[126] M. K. Leung, H.Y. Xiong, L. J. Lee, and B. J. Frey, Bioinformatics 30, 121 (2014).

[127] C. M. Bishop, Pattern Recognition and Machine Learning (Information Science and Statistics), 1st edn. 2006. corr. 2nd printing edn. Springer, New York, 2007.

[128] K. P. Murphy, Machine learning: a probabilistic perspective, MIT press, 2012.

[129] R. Sutton and A. G. Barto, Reinforcement Learning: An Introduction, A Bradford Book (1998).

[130] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, Nature 549, 195 (2017).

[131] See https://physicsml.github.io/pages/papers.html for a list of references.

[132] Y. LeCun, Y. Bengio, G. Hinton, Deep learning. Nature 521, 436 (2015).

[133] A. Morningstar and R. G. Melko, arXiv:1708.04622 (2017).

[134] J. Carrasquilla and R. G. Melko, Nat. Phys. http://dx.doi.org/10.1038/nphys4035 (2017).

[135] L. Wang, Phys. Rev. B 94, 195105 (2016).

[136] W. Hu, R. R. P. Singh, and R. T. Scalettar, Phys. Rev. E. 95, 062122 (2017).

[137] N. C. Costa, W. Hu, Z. J. Bai, R. T. Scalettar, and R. R. P. Singh, arXiv:1708.04762 (2017).

[138] S. J. Wetzel, arXiv:1703.02435 (2017).

[139] E. P. L. van Nieuwenburg, Y.-H. Liu, and S. D. Huber, Nat. Phys. http://dx.doi.org/10.1038/nphys4037 (2017).

[140] P. Broecker, F. Assaad, and S. Trebst, arXiv:1707.00663 (2017).

[141] Y.-H. Liu and E. P. L. van Nieuwenburg, arXiv:1706.08111.

[142] F. Schindler, N. Regnault, and T. Neupert, Phys. Rev. B 95, 245134 (2017).

[143] J. Venderley, V. Khemani, and E.-A. Kim, arXiv:1711.00020 (2017)

[144] L. Li, J. C. Snyder, I. M. Pelaschier, J. Huang, U.- N. Niranjan, P. Duncan, M. Rupp, K.-R. Muller, and K. Burke, International Journal of Quantum Chemistry (2015).

[145] G. Carleo and M. Troyer, Science 355, 602 (2017).

[146] G. Torlai and R. G. Melko, Phys. Rev. B 94, 165134 (2016)

[147] P. Broecker, J. Carrasquilla, R. G. Melko, and S. Trebst, arXiv:1608.07848 (2016).

[148] K. Chng, J. Carrasquilla, R. G. Melko, and E. Khatami, arXiv:1609.02552 (2016).

[149] A. Tanaka and A. Tomiya, arXiv:1609.09087 (2016).

[150] T. Ohtsuki and T. Ohtsuki, Journal of the Physical Society of Japan 85, 123706 (2016).

[151] L. Li, T. E. Baker, S. R. White, and K. Burke, Phys. Rev. B 94, 245129 (2016).

[152] D.-L. Deng, X. Li, and S. D. Sarma, arXiv:1609.09060 (2016).

[153] L. Huang and L. Wang, Phys. Rev. B 95, 035105 (2017).

[154] J. Liu, Y. Qi, Z. Y. Meng, and L. Fu, Phys. Rev. B 95, 041101 (2017).

[155] G. Torlai and R. G. Melko, Phys. Rev. Lett. 119, 030501 (2017).

[156] Y. Zhang and E.-A. Kim, arXiv:1611.01518 (2016).

[157] N. Portman and I. Tamblyn, arXiv:1611.05891 (2016).

[158] D. Crawford, A. Levit, N. Ghadermarzy, J. S. Oberoi, and P. Ronagh, arXiv:1612.05695 (2016).

[159] K.-I. Aoki and T. Kobayashi, Modern Physics Letters B 30, 1650401 (2016).

[160] D.-L. Deng, X. Li, and S. D. Sarma, arXiv:1701.04844 (2017).

[161] X. Gao and L.-M. Duan, arXiv:1701.05039 (2017).

[162] J. Chen, S. Cheng, H. Xie, L. Wang, and T. Xiang, arXiv:1701.04831 (2017).

[163] Y. Huang and J. E. Moore, arXiv:1701.06246 (2017).

[164] G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko, and G. Carleo, Nat. Phys. https://doi.org/10.1038/s41567-018-0048-5 (2018).

[165] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, Journal of Machine Learning Research 12, 2825 (2011).

[166] J. Platt, in Advances in Kernel Methods - Support Vector Learning (MIT Press, 1998).

[167] L. Onsager, Phys. Rev. 65, 117 (1944).

[168] M. Newman and G. Barkema, Monte Carlo Methods in Statistical Physics (Oxford University Press, 1999).

[169] M. J. Beach, A. Golubeva, and R. G. Melko, Phys. Rev. B 97, 045207 (2018).