# Integrability breaking and thermalization

**Federica M. Surace**[1⋆]

Department of Physics and Institute for Quantum Information and Matter, California Institute of Technology, Pasadena, California 91125, USA

⋆ fsurace@caltech.edu

**May 2025. Alert! These notes are still under construction. There might be typos, mistakes, incomplete results, missing references. Use at your own risk.**

## Abstract

**In these lectures, I will explore the concept of thermalization in isolated quantum many-body systems, and show how the eigenstate thermalization hypothesis is a useful ansatz for predicting many properties of thermal equilibrium. I will argue that conventional thermalization paradigms fail for integrable systems, and introduce a generalized notion of equilibration. I will then examine how small integrability-breaking perturbations give rise to distinct thermalization timescales, highlighting the role of "weak" integrability-breaking perturbations. Finally, I will discuss their implications for thermalization dynamics.**

## Contents

# 1 Introduction

In many traditional settings, such as electrons in solids, quantum many-body systems are coupled to an environment, which quickly drives the system to thermal equilibrium. For this reason, studies of quantum many-body physics historically focused on equilibrium properties, either at zero or finite temperature.

This picture has changed dramatically with the development of highly controlled experimental platforms like ultracold atoms, trapped ions, and superconducting qubits. These systems are nearly isolated from their environment and evolve under unitary dynamics for surprisingly long times. As a result, they provide a unique opportunity to study the non-equilibrium dynamics of many-body systems in real time.

Fundamental question is how to reconcile the isolated unitary dynamics of a quantum system with the idea that a system with many-particle should thermalize. Thermalization of isolated quantum many-body systems is a fascinating research topic that has been investigate now for two decades and is a fundamental question of statistical mechanics.

But understanding if a system reaches thermal equilibrium is not the end of the story. A key insight that has emerged is the importance of timescales. Even in systems that eventually thermalize, this process can take a long time. On intermediate timescales, systems may exhibit slow relaxation, retain memory of initial conditions, or show signatures of approximate conservation laws. Understanding how these timescales arise, how they depend on microscopic parameters, and how they relate to the eventual onset of thermal behavior has become a central goal — not only of theoretical interest, but also directly relevant to current experiments in quantum simulation.

# 2 Recommended resources

For those interested in deeper exploration, the following resources are highly recommended. Much of the material in these lecture notes draws inspiration from them.

**Many-body dynamics and thermalization**

- Lecture notes of the course "Many-body quantum dynamics" held in Dresden by Marin Bukov, Pieter Claeys, Johannes Mitscherling, and Roderich Moessner.
- Lecture notes *Chaotic dynamics: the Eigenstate Thermalization Hypothesis and beyond* by Silvia Pappalardi, from the school *Lectures on Statistical Field Theories, GGI, Florence*.
- Review article *From quantum chaos and eigenstate thermalization to statistical mechanics and thermodynamics*, by Luca D'Alessio, Yariv Kafri, Anatoli Polkovnikov, and Marcos Rigol, Advances in Physics, 65(3), 239-362.

**Dynamics in integrable systems**

- Review article *Quench dynamics and relaxation in isolated integrable quantum spin chains* by Fabian Essler, and Maurizio Fagotti, J. Stat. Mech. (2016) 064002.
- Lecture notes *Entanglement spreading in non-equilibrium integrable systems*, by Pasquale Calabrese, SciPost Phys. Lect. Notes 20 (2020), from the Les Houches Summer School on *Integrability in Atomic and Condensed Matter Physics* in summer 2018

**Prethermalization**

- Review article *Thermalization and prethermalization in isolated quantum systems: a theoretical overview*, by Takashi Mori, Tatsuhiko N. Ikeda, Eriko Kaminishi, and Masahito Ueda, J. Phys. B 51, 112001 (2018).

**Weak integrability breaking**

- *Current operators in integrable spin chains: lessons from long range deformations* by Balázs Pozsgay, SciPost Phys. 8, 016 (2020).
- *Weak integrability breaking perturbations of integrable models*, by Federica Maria Surace, and Olexei Motrunich, Phys. Rev. Research 5, 043019 (2023).

# 3 Equilibration and thermalization

We are familiar with the fact that systems tend to **equilibrium**. Equilibrium states are described by a few quantities (e.g., temperature, chemical potential). In time evolution, all the other microscopic information about the initial state is lost. But microscopically there is no arrow of time. While this is an interesting question both in classical and quantum many-body systems, we will focus on the latter.

## 3.1 Dynamics in quantum many-body systems

In these lectures, we focus on isolated quantum many-body systems, whose evolution is given by a time-independent Hamiltonian $\hat{H}$. We assume that $\hat{H}$ is a sum of local terms on a $D$-dimensional lattice. We call $N$ the number of sites, and $\mathcal{D}_N$, is the Hilbert space dimension, which grows exponentially with $N$.

In quantum mechanics, the time evolution of an isolated system is described by the Schroedinger equation (we set $\hbar = 1$)

$$i\partial_t |\psi(t)\rangle = \hat{H} |\psi(t)\rangle \,. \tag{1}$$

We can solve it, to get

$$|\psi(t)\rangle = e^{-i\hat{H}t}\,|\psi_0\rangle = \sum_n C_n e^{-iE_n t}\,|E_n\rangle\,, \tag{2}$$

where we used the spectral decomposition of the Hamiltonian $\hat{H} = \sum_n E_n\,|E_n\rangle\,\langle E_n|$ and we defined $C_n \equiv \langle E_n|\psi_0\rangle$.

The evolution is unitary, there is no loss of information. From the state at time $t$ we can always recover the state at time 0 by evolving backwards. How can we talk about the thermalization in this case?

Rather than considering $|\psi(t)\rangle$ we should focus on **local observables**. Let us consider the expectation value of a local observable $\hat{O}$. If the system equilibrates, then the expectation value of $\hat{O}$ should reach a stationary value

$$\langle\hat{O}(t)\rangle \equiv \langle\psi(t)|\,\hat{O}\,|\psi(t)\rangle \xrightarrow[t\to\infty]{} \langle\hat{O}\rangle_{\text{eq}}\,. \tag{3}$$

Using Eq. (2) we can write the time evolved expectation value as

$$\langle\hat{O}(t)\rangle = \sum_n |C_n|^2 O_{nn} + \sum_{n\neq m} C_n^* C_m O_{nm} e^{-i(E_m-E_n)t}, \tag{4}$$

where we separated the diagonal and off-diagonal contribution in the Hamiltonian eigenbasis. Now we make the following observation: if Eq. (3) is true, then the time-averaged expectation value of $\langle\hat{O}(t)\rangle$ should also converge to $\langle\hat{O}\rangle_{\text{eq}}$, i.e.,

$$\frac{1}{T}\int_0^T dt\,\langle\hat{O}(t)\rangle \xrightarrow[T\to\infty]{} \langle\hat{O}\rangle_{\text{eq}}\,. \tag{5}$$

We can evaluate the left-hand side from Eq. (4). Assuming that there are no degeneracies in the spectrum[1], we get that the off-diagonal terms average to zero and we are left with the diagonal part only:

$$\frac{1}{T}\int_0^T dt\,\langle\hat{O}(t)\rangle \xrightarrow[T\to\infty]{} \sum_n |C_n|^2 O_{nn}. \tag{6}$$

The result in Eq. (6) can be conveniently expressed by defining the **diagonal ensemble**

$$\hat{\rho}_{DE} = \sum_n |C_n|^2\,|E_n\rangle\,\langle E_n|\,. \tag{7}$$

Then

$$\langle\hat{O}\rangle_{\text{eq}} = \sum_n |C_n|^2 O_{nn} = \text{Tr}[\hat{O}\hat{\rho}_{DE}]. \tag{8}$$

We have a few observations to make on this result. First of all, we had to assume that the system equilibrates, i.e., that the limit in Eq. (3) exists. In fact, from the expression Eq. (6) we could only argue that the *time average* of $\langle\hat{O}(t)\rangle$ has a well-defined limit at long times. We still have no clue about the fluctuations of $\langle\hat{O}(t)\rangle$ over time: it may be highly oscillating, and the limit Eq. (3) might not exist. Secondly, to argue that the limit of the time averaged expectation value exists we had to average out all the phases $e^{-i(E_n-E_m)t}$. For this to hold, we need to wait for a time $T \gtrsim \min_{(E_n,E_m)}|E_n - E_m|$. This time is exponentially long in the system size! This doesn't make sense. We know that thermalization shouldn't take so long. Finally, the diagonal ensemble depends on the population of each of the exponentially many

[1]The case of degeneracies can be easily dealt with: for each degenerate subspace, we choose the projection of the initial state on the subspace as one of the basis states, such that $C_n$ is non-zero for a single eigenstate with energy $E_n$.

eigenstates. These are definitely too many quantities. We want something that depends only on a few properties of the initial state. In particular, if energy is the only conserved quantity, we expect that thermal equilibrium will correspond to the Gibbs ensemble that is specified only by the inverse temperature $\beta$

$$\hat{\rho}_{\text{Gibbs}} = \frac{e^{-\beta \hat{H}}}{\text{Tr}[e^{-\beta \hat{H}}]}. \tag{9}$$

All these will be clear in the next section. We will show how the eigenstate thermalization hypothesis will solve them all. For the moment let us stress one more point about the interpretation of thermalization for local observables.

If

$$\langle \hat{O}(t) \rangle \xrightarrow[t\to\infty]{} \text{Tr}[\hat{\rho}_{\text{eq}} \hat{O}] \tag{10}$$

for all observables with support in a subsystem $A$, then

$$\hat{\rho}_A(t) \equiv \text{Tr}_{\overline{A}}[|\psi(t)\rangle \langle\psi(t)|] \xrightarrow[t\to\infty]{} \text{Tr}_{\overline{A}}[\hat{\rho}_{\text{eq}}]. \tag{11}$$

This is what we mean by equilibration: of course $|\psi(t)\rangle \langle\psi(t)|$ is a pure state, it cannot become a mixed state like $\hat{\rho}_{\text{Gibbs}}$, but when we trace over $\overline{A}$, the limit is well-defined. In the rest of the notes, we will often be sloppy and say that $|\psi(t)\rangle \langle\psi(t)| \xrightarrow[t\to\infty]{} \hat{\rho}_{\text{Gibbs}}$, but what we will actually mean is Eq (11). This is also a crucial point underlying the equivalence of different thermodynamic ensembles: as we will show later, we could choose to describe $\hat{\rho}_{\text{eq}}$ as a microcanonical or a canonical ensemble, and we will get the same result for $\text{Tr}_{\overline{A}}[\rho_{\text{eq}}]$.

The intuitive picture for Eq. (11) is that thermalization occurs because $\overline{A}$ acts as a bath for $A$. We should also keep in mind that when we talk about thermalization we always need to take the thermodynamic limit $V_{\overline{A}} \gg V_A$, and we have to take this limit *before* taking the limit $t \to \infty$.

### 3.2 Eigenstate thermalization hypothesis

The general idea behind ETH is to bring statistical assumptions into the quantum many-body problem. Instead of describing individual eigenstates with full microscopic detail, one introduces randomness into the structure of eigenstates and matrix elements of observables. This perspective dates back to Wigner's work on nuclear spectra, where he proposed modeling complex Hamiltonians using random matrices. Berry later suggested that high-energy eigenstates of classically chaotic systems resemble random vectors, an idea now known as the Berry conjecture.

Building on these insights, the eigenstate thermalization hypothesis provides a mechanism for the emergence of thermal behavior in isolated quantum systems. It proposes a specific structure for the matrix elements of few-body observables $\hat{O}$ in the energy eigenbasis of a chaotic many-body Hamiltonian:

$$O_{nm} = \overline{O}(E)\delta_{nm} + e^{-S(E)/2} f_O(E, \omega) R_{mn} \tag{12}$$

where:

- $E = \frac{E_n + E_m}{2}$ is the average energy,
- $\omega = E_n - E_m$ is the energy difference,
- $\overline{O}(E)$ is a smooth function that captures the microcanonical expectation value of $\hat{O}$ at energy $E$,

- $S(E)$ is the thermodynamic entropy at energy $E$, defined as $S(E) = \log(\rho(E))$ where $\rho(E)$ is the density of states (i.e., $\rho(E)dE$ is the number of energy eigenstates in the interval $[E, E+dE]$),
- $f_O(E,\omega)$ is a smooth function of its arguments that controls the decay of off-diagonal terms,
- $R_{nm}$ is a random variable with zero mean and unit variance, encoding statistical fluctuations.

We now explore what the Eigenstate Thermalization Hypothesis reveals about thermalization following a quantum quench, the equivalence of statistical ensembles, and the behavior of dynamical correlation functions.

### 3.2.1 Diagonal part: Quench dynamics

The diagonal part of the ETH ansatz, $O_{nn} = \overline{O}(E_n)$. captures the thermodynamic expectation values of observables in individual eigenstates. This is at the heart of thermalization: the long-time behavior of observables after a quantum quench is controlled by the diagonal matrix elements in the energy eigenbasis (Eq. (8)).

As we discussed in Sec. 3.1, in unitary dynamics following a quench, the system evolves from an initial state $|\psi_0\rangle$, and long-time observables equilibrate to the so-called diagonal ensemble.

$$\langle \hat{O} \rangle_{eq} = \sum_n |C_n|^2 O_{nn} \to \sum_n |C_n|^2 \overline{O}(E_n) \tag{13}$$

We will assume that the energy distribution of the initial states, encoded in the overlaps $|C_n|^2$, is sufficiently narrow around $\langle E \rangle = \langle \psi_0 | \hat{H} | \psi_0 \rangle$ and one can expand $\overline{O}(E_n)$ around $\langle E \rangle$ (we will check this assumption later):

$$\overline{O}(E_n) \approx \overline{O}(\langle E \rangle) + \overline{O}'(\langle E \rangle)(E_n - \langle E \rangle) + \frac{1}{2}\overline{O}''(\langle E \rangle)(E_n - \langle E \rangle)^2. \tag{14}$$

Inserting this into the expression for $\langle \hat{O} \rangle_{\text{eq}}$, and using the normalization and definitions of the moments:

$$\langle \hat{O} \rangle_{eq} \approx \overline{O}(\langle E \rangle) \sum_n |C_n|^2 + \overline{O}'(\langle E \rangle) \sum_n |C_n|^2 (E_n - \langle E \rangle) + \frac{1}{2}\overline{O}''(\langle E \rangle) \sum_n |C_n|^2 (E_n - \langle E \rangle)^2 \tag{15}$$

$$= \overline{O}(\langle E \rangle) + \frac{1}{2}\overline{O}''(\langle E \rangle)\, \delta E^2 \tag{16}$$

Let us now analize when this expansion around $\langle E \rangle$ is justified. We can write the Hamiltonian as $\hat{H} = \sum_j \hat{h}_j$, where $\hat{h}_j$ is the energy density. Then

$$\delta E^2 = \sum_{ij} (\langle \hat{h}_i \hat{h}_j \rangle - \langle \hat{h}_i \rangle \langle \hat{h}_j \rangle) = \sum_i \sum_r \langle \hat{h}_i \hat{h}_{i+r} \rangle_c \,. \tag{17}$$

A reasonable assumption in many cases is that the initial state has exponentially decaying connected correlations. If the connected correlator $\langle \hat{h}_i \hat{h}_{i+r} \rangle_c$ decays exponentially with $r$, then $\sum_r \langle \hat{h}_i \hat{h}_{i+r} \rangle_c$ is finite in the thermodynamic limit, and we are left with an extensive quantity. Therefore the energy variance of the initial state scales with the system size $N$ as

$$\delta E^2 = \langle \psi_0 | (\hat{H} - \langle E \rangle)^2 | \psi_0 \rangle \propto N, \tag{18}$$

where $\langle E\rangle = \langle\psi_0|\hat{H}|\psi_0\rangle$ is the average energy. Under this assumption, the term $O''(\langle E\rangle)\delta E^2$ in Eq. (16) is proportional to $N^{-1}$ because each derivate with respect to $E$ (which is extensive) gives a factor $N^{-1}$. Therefore, the expansion is justified and we get

$$\langle\hat{O}\rangle_{eq} \approx \overline{O}(\langle E\rangle), \tag{19}$$

up to subleading corrections of order $N^{-1}$. This result shows that the equilibrium value (obtained from the diagonal ensemble) is equal, in the thermodynamic limit, to the microcanonical expectation value at energy $\langle E\rangle$. This solves one of the issues of the diagonal ensemble that we mentioned above: according to the diagonal ensemble, the equilibrium state depends on the initial state through an exponentially many quantities (the overlaps $C_n$ with the energy eigenstates). What this result shows, is that we do not need so many quantities. The equilibrium state depends on the initial state only through a single quantity, the initial energy $\langle E\rangle$.

### 3.2.2 Diagonal part: Canonical expectation values

In the canonical ensemble at inverse temperature $\beta$, the expectation value of an observable is given by

$$\langle\hat{O}\rangle_\beta = \frac{1}{Z}\sum_n e^{-\beta E_n} O_{nn}, \tag{20}$$

where $Z = \sum_n e^{-\beta E_n}$ is the partition function. Using ETH, we can replace $O_{nn} \to \overline{O}(E_n)$, and in the thermodynamic limit, the sum over eigenstates becomes an integral over energies:

$$\sum_n \delta(E - E_n) \to \rho(E) = e^{S(E)}, \tag{21}$$

where $S(E)$ is the thermodynamic entropy and $\rho(E)$ is the density of states.

Hence, the expectation value becomes

$$\langle\hat{O}\rangle_\beta = \frac{1}{Z}\int dE \left(\sum_n \delta(E - E_n)\right) e^{-\beta E}\overline{O}(E) \to \frac{1}{Z}\int dE\, e^{S(E)-\beta E}\,\overline{O}(E). \tag{22}$$

The integral is dominated by a saddle point at energy $E_\beta$ defined by $S'(E_\beta) = \beta$. At leading order,

$$\langle\hat{O}\rangle_\beta \approx \overline{O}(E_\beta), \tag{23}$$

i.e., the thermal expectation value equals the ETH function evaluated at the saddle-point energy. We thus showed that the canonical expectation value is equivalent, in the thermodynamic limit, to the microcanonical expectation value at energy $E_\beta$.

This result also gives us a prescription for finding the inverse temperature $\beta$ of the Gibbs state describing the equilibrium state reached under evolution from an initial state of energy $\langle E\rangle$. Combining this result with the final result of the previous section, we get $\beta = S'(\langle E\rangle)$. An equivalent prescription, which is often useful when we do not have access to the thermodynamic entropy, is to find $\beta(\langle E\rangle)$ such that the energy (or more precisely, the energy density) of the corresponding Gibbs state matches the energy (density) of the initial state, i.e.,

$$\frac{1}{Z}\mathrm{Tr}[e^{-\beta(\langle E\rangle)\hat{H}}\hat{H}] = \langle E\rangle\,. \tag{24}$$

### 3.2.3 Off-diagonal part: time fluctuations

So far, we have focused on the diagonal part of the ETH ansatz. We now turn to the off-diagonal matrix elements and explore their implications for the time dependence of observables.

We already showed that the long-time average of a quantum expectation value is governed by the diagonal ETH term:

$$\frac{1}{T}\int_0^T dt\,\langle\hat{O}(t)\rangle \xrightarrow[T\to\infty]{} \sum_n |C_n|^2 O_{nn} \approx \overline{O}(\langle E\rangle), \tag{25}$$

where we have again neglected corrections subleading in the system size.

To understand whether the system indeed equilibrates, we want to know how the system fluctuates around this average at late times. We hence compute the variance in time:

$$\delta O_t^2 = \lim_{T\to\infty}\frac{1}{T}\int_0^T dt\,[\langle\hat{O}(t)\rangle - \overline{O}(\langle E\rangle)]^2. \tag{26}$$

Plugging in the spectral decomposition of $\hat{O}(t)$, and using the fact that the diagonal part gives a time-independent contribution, we find that only the off-diagonal terms contribute to fluctuations:

$$\delta O_t^2 = \lim_{T\to\infty}\frac{1}{T}\int_0^T dt\,\left|\sum_{n\neq m} C_n^* C_m O_{nm} e^{i(E_n-E_m)t}\right|^2. \tag{27}$$

Squaring the sum gives rise to a double sum over indices $(n_1,m_1),(n_2,m_2)$, with an oscillating phase:

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T dt, e^{i(E_{n_1}-E_{m_1}-E_{n_2}+E_{m_2})t} = \begin{cases}1, & \text{if } E_{n_1}-E_{m_1}=E_{n_2}-E_{m_2}\\ 0, & \text{otherwise.}\end{cases} \tag{28}$$

Assuming that the only resonant contributions arise when $n_1 = n_2$ and $m_1 = m_2$, the expression simplifies to

$$\delta O_t^2 = \sum_{n\neq m} |C_n|^2 |C_m|^2 |O_{nm}|^2. \tag{29}$$

This sum is suppressed by the typical size of off-diagonal matrix elements. According to ETH, these matrix elements scale as

$$|O_{nm}|^2 \sim e^{-S(\langle E\rangle)},$$

where $S(\langle E\rangle)$ is the thermodynamic entropy at the mean energy. This gives

$$\delta O_t^2 \lesssim \max_{n\neq m}|O_{nm}|^2 \sim e^{-S(\langle E\rangle)}, \tag{30}$$

which vanishes exponentially in system size. In large systems, time fluctuations become negligible: $\langle\hat{O}(t)\rangle$ is close to its equilibrium value at long times, and we do not need to average over time to find an agreement with thermal expectation values.

### 3.2.4 Off-diagonal part: dynamical correlation functions

Off-diagonal matrix elements also control the behavior of time-dependent correlation functions in equilibrium. One of the simplest examples is the two-point function

$$\kappa_2(t) = \langle\hat{O}(t)\hat{O}\rangle_\beta - \langle\hat{O}\rangle_\beta^2, \tag{31}$$

which measures fluctuations around thermal expectation values.

Inserting a complete set of energy eigenstates and using the ETH ansatz, one finds that the time dependence is governed by the structure function $|f_O(E,\omega)|^2$, which controls the typical size of off-diagonal matrix elements at a given energy difference $\omega = E_n - E_m$:

$$\kappa_2(t) = \int d\omega\, e^{-\beta\omega/2} e^{i\omega t} |f_O(E_\beta, \omega)|^2 \tag{32}$$

The function $f_O(E,\omega)$ thus encodes the full spectral and temporal structure of the correlation. The shape of this function in $\omega$ determines how quickly the correlation function decays. This quantity also plays an important role in the fluctuation-dissipation theorem.

**Exercise 1** *The fluctuation-dissipation theorem relates the quantity*

$$\tilde{F}(\omega) = \frac{1}{2\pi}\int dt\, e^{i\omega t} F(t), \qquad F(t) = \frac{1}{2}\langle\{\hat{O}(t), \hat{O}\}\rangle_\beta - \langle\hat{O}\rangle_\beta^2, \tag{33}$$

*which quantifies the thermal fluctuations in the system and*

$$\tilde{\rho}(\omega) = \frac{1}{2\pi}\int dt\, e^{i\omega t} \rho(t), \qquad \rho(t) = \langle[\hat{O}(t), \hat{O}]\rangle_\beta, \tag{34}$$

*which appears in linear-response theory and quantifies the dissipative response to at external perturbation at frequency $\omega$. The theorem states that*

$$\tilde{F}(\omega) = \frac{1}{2}\coth\left(\frac{\beta\omega}{2}\right)\tilde{\rho}(\omega). \tag{35}$$

*Use Eq. (32) to prove the fluctuation-dissipation theorem.*

### 3.3 Dynamics in integrable systems and the generalized Gibbs ensemble

A broad class of quantum systems that do not obey the eigenstate thermalization hypothesis (ETH) are *integrable systems*. These systems possess an extensive number of local conserved quantities, which strongly constrain their dynamics and prevent thermalization in the conventional sense.

#### Example: XX Spin Chain

A simple example of an integrable system is the XX spin chain, described by the Hamiltonian

$$\hat{H} = \sum_j \left(\hat{\sigma}_j^x \hat{\sigma}_{j+1}^x + \hat{\sigma}_j^y \hat{\sigma}_{j+1}^y\right), \tag{36}$$

which preserves the total magnetization in the $z$-direction.

This model can be mapped to a system of free spinless fermions via the Jordan-Wigner transformation:

$$\hat{\sigma}_j^+ = \exp\left(i\pi \sum_{l<j} \hat{c}_l^\dagger \hat{c}_l\right)\hat{c}_j^\dagger, \tag{37}$$

$$\hat{\sigma}_j^- = \exp\left(-i\pi \sum_{l<j} \hat{c}_l^\dagger \hat{c}_l\right)\hat{c}_j, \tag{38}$$

$$\hat{\sigma}_j^z = 2\hat{c}_j^\dagger \hat{c}_j - 1. \tag{39}$$

Using this transformation, the Hamiltonian becomes

$$\hat{H} = -\sum_j \left( \hat{c}_j^\dagger \hat{c}_{j+1} + \hat{c}_{j+1}^\dagger \hat{c}_j \right), \tag{40}$$

which describes free fermions hopping on a one-dimensional lattice.

We can diagonalize this model in momentum space by introducing Fourier-transformed operators:

$$\hat{c}_j = \frac{1}{\sqrt{L}} \sum_k e^{ikj} \hat{c}_k, \qquad \hat{c}_k = \frac{1}{\sqrt{L}} \sum_j e^{-ikj} \hat{c}_j, \tag{41}$$

which yields the diagonal form

$$\hat{H} = -2 \sum_k \cos(k)\, \hat{c}_k^\dagger \hat{c}_k. \tag{42}$$

The operators $\hat{n}_k = \hat{c}_k^\dagger \hat{c}_k$ are conserved quantities: the momentum occupation numbers are constants of motion. While the $\hat{n}_k$'s are non-local in real space, it is sometimes more convenient to work with a set of **local conserved charges**, i.e., operators that can be written as a sum over space of local densities:

$$\hat{Q}_m = \sum_j \hat{q}_j^{(m)}, \tag{43}$$

with $\hat{q}_j^{(m)}$ acting on $m+1$ adjacent sites These can be constructed as:

$$\hat{Q}'_m = \sum_k 2\cos(mk)\, \hat{n}_k = \sum_j (\hat{c}_j^\dagger \hat{c}_{j+m} + \hat{c}_{j+m}^\dagger \hat{c}_j). \tag{44}$$

$$\hat{Q}''_m = \sum_k 2\sin(mk)\, \hat{n}_k = \sum_j i(\hat{c}_j^\dagger \hat{c}_{j+m} - \hat{c}_{j+m}^\dagger \hat{c}_j). \tag{45}$$

Thus, the local conserved quantities are obtained by taking cosine- and sine-weighted combinations of the momentum occupation numbers. Both representations – in terms of the non-local mode occupations $\hat{n}_k$, or a set of local charges $\hat{Q}_m$ – are complete and equivalent, and uniquely characterize the steady state of local observables and define the GGE for the XX chain.

**Interacting Integrable Systems**

The XX spin chain is a special case of an integrable system that can be mapped to free fermions. This makes its dynamics and thermodynamics particularly transparent. Other integrable systems, such as the XXZ chain,

$$\hat{H} = \sum_j \left( \hat{\sigma}_j^x \hat{\sigma}_{j+1}^x + \hat{\sigma}_j^y \hat{\sigma}_{j+1}^y + \Delta \hat{\sigma}_j^z \hat{\sigma}_{j+1}^z \right), \tag{46}$$

are not mappable to free fermions (except at special points like $\Delta = 0$), but can be solved using the **Bethe Ansatz**, a method first introduced by Hans Bethe in 1931. The central idea is to write the many-body eigenstates as superpositions of plane waves, whose momenta are not free but are determined by a set of coupled nonlinear equations—the *Bethe equations*. The Bethe Ansatz allows us to interpret the eigenstates as consisting of **quasiparticles**—stable collective excitations labeled by rapidities $\lambda$. Unlike free particles, these quasiparticles interact through two-body scattering, but integrability ensures that the scattering is elastic—quasiparticles retain their identity and momenta after collisions, with no particle production or reshuffling. Moreover, all many-body scattering processes factorize into a sequence of two-body scatterings.

In these systems, as already illustrated in the XX chain, integrability is associated with the existence of an infinite set of conserved charges $\hat{Q}_\alpha$ that commute with the Hamiltonian and with each other: $[\hat{H}, \hat{Q}_\alpha] = 0, \quad [\hat{Q}_\alpha, \hat{Q}_\beta] = 0$. These include not only energy and momentum, but also other local (or quasi-local), quantities that restrict the accessible phase space.

### Generalized Gibbs Ensemble

Despite the absence of thermalization in the ETH sense, integrable systems do *equilibrate*: local observables relax to steady values at long times. However, these values are not described by the standard Gibbs ensemble, since additional conservation laws must be taken into account.

The appropriate statistical ensemble is the **generalized Gibbs ensemble** (GGE), defined by

$$\hat{\rho}_{\mathrm{GGE}} = \frac{e^{-\sum_\alpha \mu_\alpha \hat{Q}_\alpha}}{\mathrm{Tr}\left[e^{-\sum_\alpha \mu_\alpha \hat{Q}_\alpha}\right]}, \tag{47}$$

where $\mu_\alpha$ are generalized chemical potentials fixed by the initial state:

$$\mathrm{Tr}(\hat{\rho}_{\mathrm{GGE}} \hat{Q}_\alpha) = \langle \psi_0 | \hat{Q}_\alpha | \psi_0 \rangle \,.$$

The GGE captures the steady-state values of local observables after a quantum quench in integrable systems. It reduces to the standard canonical ensemble when only energy is conserved. The success of the GGE relies on the assumption that, in the thermodynamic limit, local observables are completely determined by the set of conserved charges.

This framework has been tested in a variety of systems, both theoretically and experimentally—particularly in cold atom setups where integrability is approximately realized.

### Currents and generalized currents

Integrable systems have an extensive set of conserved charges $\hat{Q}_\alpha$. Each of these can be written as the sum over lattice sites of a charge density operator $\hat{q}_{\alpha,j}$ that has finite support[2] around site $j$:

$$\hat{Q}_\alpha = \sum_j \hat{q}_{\alpha,j}. \tag{48}$$

The conservation of the total charge $\hat{Q}_\alpha$ implies the existence of a current operator $\hat{J}_{\alpha,j}$ that satisfies the continuity equation:

$$\frac{d}{dt}\hat{q}_{\alpha,j} = i[\hat{H}_0, \hat{q}_{\alpha,j}] = \hat{J}_{\alpha,j} - \hat{J}_{\alpha,j+1}. \tag{49}$$

Here $\hat{J}_{\alpha,j}$ is a local operator, with finite support around site $j$.

The existence of these currents was inferred from the commutation $[\hat{H}, \hat{Q}_\alpha] = 0$. In integrable models, the conserved quantities don't just commute with $\hat{H}$, they also mutually commute. Therefore, we have a continuity equation for each pair of charges $(\hat{Q}_\alpha, \hat{Q}_\beta)$ and we can define the generalized currents:

$$i[\hat{Q}_\beta, \hat{q}_{\alpha,j}] = \hat{J}_{\alpha\beta,j} - \hat{J}_{\alpha\beta,j+1}. \tag{50}$$

These generalized currents will later be useful when we will consider integrability breaking perturbations.

[2] Additional conserved quantities that are quasilocal operators can also be found in some cases. We will not consider them here.

# 4 Integrability breaking and prethermalization

We've seen that integrable systems behave very differently from ergodic ones when it comes to thermalization, preserving an extensive number of local conserved quantities. Integrability is, however, a fine-tuned property: even a small perturbation that breaks integrability should eventually drive the system toward standard thermal behavior. The big question is *how* this crossover happens. In this section, we aim to understand this mechanism.

We consider a system governed by a Hamiltonian of the form

$$\hat{H} = \hat{H}_0 + \lambda \hat{V}, \tag{51}$$

where $\hat{H}_0$ is integrable, $\lambda \ll 1$ is a small parameter, and $\hat{V}$ is a perturbation that breaks integrability. Our discussion also applies when $\hat{H}_0$ possesses some conservation laws that are broken by $\hat{V}$, even if $\hat{H}_0$ is not fully integrable. We assume that $\hat{V}$ is translationally invariant, local, and extensive.

For times much shorter than $\lambda^{-1}$, the dynamics are approximately governed by $\hat{H}_0$. In fact, one can rigorously bound the difference between expectation values evolved with $\hat{H}$ and $\hat{H}_0$ as (see Appendix C for the proof):

$$\left| \langle \psi_0 | e^{i\hat{H}t} \hat{O} e^{-i\hat{H}t} | \psi_0 \rangle - \langle \psi_0 | e^{i\hat{H}_0 t} \hat{O} e^{-i\hat{H}_0 t} | \psi_0 \rangle \right| \leq \lambda C t, \tag{52}$$

where $C$ is a constant independent of system size (under standard assumptions).

Suppose that evolution under $\hat{H}_0$ drives the system to a Generalized Gibbs Ensemble (GGE) on a timescale $\tau_0$. Then, as long as $\lambda^{-1} \gg \tau_0$, the system first relaxes to the GGE — as if it were evolving under $\hat{H}_0$ — and only later does the integrability-breaking perturbation become effective. Eventually, the system crosses over to a Gibbs ensemble consistent with $\hat{H}$. This two-stage relaxation process is known as **prethermalization**.

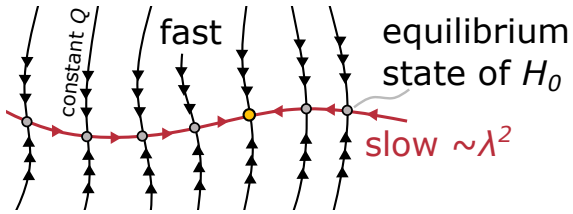


Figure 1: The dynamics induced by $\hat{H}_0$ is constrained to a subspace of quantum states with constant $\langle \hat{Q} \rangle$, and proceeds to the equilibrium state of $\hat{H}_0$ in such subspace. The effect of a small perturbation $\lambda \hat{V}$ that breaks the conservation law $\hat{Q}$ is to induce a slow drift between equilibrium states of $\hat{H}_0$ with different $\langle \hat{Q} \rangle$, eventually taking it to the equilibrium state of $\hat{H}_0 + \lambda \hat{V}$.

The idea is sketched in Figure 1: the fast dynamics under $\hat{H}_0$ project the system onto a manifold where the conserved quantities $\hat{Q}_\alpha$ are (approximately) constant. The slow drift induced by $\hat{V}$ gradually erodes these conservation laws, eventually leading to thermalization.

## 4.1 Kinetic equation

This late-time evolution can be understood as a kinetic process in the space of GGE states. We assume that at late times, the system remains in a local GGE form, but with time-dependent Lagrange multipliers:

$$\hat{\rho}(t) \approx \frac{e^{-\mu_\alpha(t)\hat{Q}_\alpha}}{\mathrm{Tr}\left[e^{-\mu_\alpha(t)\hat{Q}_\alpha}\right]}. \tag{53}$$

These Lagrange multipliers evolve due to the weak violation of the conservation laws. We will now show that this evolution can be captured by a simple master equation, where transition rates are obtained using Fermi's golden rule.

We assume a separation of timescales:

- $\tau_0$: relaxation time under $H_0$ to the GGE.
- $\lambda^{-1}$: timescale at which $\hat{V}$ becomes effective.
- $\Delta t$: coarse-graining time such that $\tau_0 \ll \Delta t \ll \lambda^{-1}$.

We also assume that at these late times the density matrix is approximately diagonal in the eigenbasis of $\hat{H}_0$:

$$\hat{\rho}(t) \approx \sum_n P_n(t) \left|E_n^{(0)}\right\rangle \left\langle E_n^{(0)}\right|, \tag{54}$$

where $\hat{H}_0 \left|E_n^{(0)}\right\rangle = E_n^{(0)} \left|E_n^{(0)}\right\rangle$. This is a rewriting of the GGE. Once again, this assumption has to be understood at the level of local observables: the state $\hat{\rho}(t)$ is not equal the time evolved state $e^{-i\hat{H}t} \left|\psi_0\right\rangle \left\langle\psi_0\right| e^{i\hat{H}t}$, they give same expectation values of local observables.

The evolution of the ensemble $\rho(t)$ is simply described in terms of the evolution of the populations:

$$\frac{P_n(t+\Delta t) - P_n(t)}{\Delta t} = \sum_{m \neq n} \left[W_{m\to n} P_m(t) - W_{n\to m} P_n(t)\right], \tag{55}$$

where the transition rates are given by Fermi's golden rule:

$$W_{n\to m} = 2\pi\lambda^2 |V_{nm}|^2 \delta(\omega_{nm}), \tag{56}$$

with $\omega_{nm} = E_n^{(0)} - E_m^{(0)}$.

This lets us derive the dynamics of conserved quantities $\hat{Q}$:

$$\frac{\langle\hat{Q}(t+\Delta t)\rangle - \langle\hat{Q}(t)\rangle}{\Delta t} = 2\pi\lambda^2 \sum_{n,m} (Q_{nn} - Q_{mm}) P_m(t) |V_{nm}|^2 \delta(\omega_{nm}), \tag{57}$$

which describes a slow drift in the conserved quantities induced by $\hat{V}$, mediated by near-resonant transitions.

Note that this discussion, that is here presented in the context of integrable systems with an extensive number of conserved quantities, can be similarly applied to the case of a finite number of conserved quantities, that are broken by the perturbation $\hat{V}$. See Ref. [1] for a more complete discussion of the problem.

## 5 Weak integrability breaking

From the kinetic equation, we see that relaxation to the Gibbs ensemble typically occurs at a rate $\sim \lambda^2$. But there are interesting cases where thermalization is even slower — with rates $\sim \lambda^{2n}$, for some integer $n \geq 2$. We refer to such $\hat{V}$ as **weakly integrability breaking**.

The intuition is the following. Let us call $\hat{Q}^{(0)}$ conserved quantity of the unperturbed Hamiltonian $[\hat{H}_0, \hat{Q}^{(0)}] = 0$. Suppose we can construct a local operator $\hat{Q}^{(1)}$ such that:

$$[\hat{Q}^{(0)} + \lambda\hat{Q}^{(1)}, \hat{H}_0 + \lambda\hat{V}] = O(\lambda^2), \tag{58}$$

i.e., $\hat{Q}^{(0)} + \lambda\hat{Q}^{(1)}$ is conserved up to $O(\lambda^2)$. This implies that integrability is only effectively broken at second order in $\lambda$, and thermalization occurs on a timescale $\sim \lambda^{-4}$, not $\lambda^{-2}$.

The condition above is equivalent to:

$$[\hat{Q}^{(1)}, \hat{H}_0] = -[\hat{Q}^{(0)}, \hat{V}]. \tag{59}$$

such that

$$[\hat{Q}^{(0)} + \lambda\hat{Q}^{(1)}, \hat{H}_0 + \lambda\hat{V}] = \lambda^2[\hat{Q}^{(1)}, \hat{V}]. \tag{60}$$

A natural solution arises if $\hat{V}$ can be written as

$$\hat{V} = i[\hat{X}, \hat{H}_0] + \hat{D}, \tag{61}$$

where $\hat{D}$ commutes with both $\hat{H}_0$ and $\hat{Q}^{(0)}$. Then, using the Jacobi identity, we get:

$$[\hat{Q}^{(0)}, \hat{V}] = -[i[\hat{X}, \hat{Q}^{(0)}], \hat{H}_0], \tag{62}$$

so we can define

$$\hat{Q}^{(1)} = i[\hat{X}, \hat{Q}^{(0)}], \tag{63}$$

and we're done.

### 5.1 Example

Here we consider the concrete example of the Heisenberg (or XXX) chain

$$\hat{H}_0 = \sum_j \vec{\sigma}_j \cdot \vec{\sigma}_{j+1}. \tag{64}$$

This is the integrable model introduced in Eq. (46) with $\Delta = 1$. The first few conserved quantities are

$$\hat{Q}_2^{(0)} = \frac{1}{2}\sum_j \vec{\sigma}_j \cdot \vec{\sigma}_{j+1}, \tag{65}$$

$$\hat{Q}_3^{(0)} = -\frac{1}{2}\sum_j (\vec{\sigma}_j \times \vec{\sigma}_{j+1}) \cdot \vec{\sigma}_{j+2}, \tag{66}$$

$$\hat{Q}_4^{(0)} = -\sum_j [(\vec{\sigma}_j \times \vec{\sigma}_{j+1}) \times \vec{\sigma}_{j+2}] \cdot \vec{\sigma}_{j+3} + \vec{\sigma}_j \cdot \vec{\sigma}_{j+2} - 2\vec{\sigma}_j \cdot \vec{\sigma}_{j+1}. \tag{67}$$

We perturb this model with the next-nearest neighbor coupling

$$\hat{V} = \sum_j \vec{\sigma}_j \cdot \vec{\sigma}_{j+2}. \tag{68}$$

As we will prove later, this is a weak perturbation. For example, we define the following correction $\hat{Q}_3^{(1)}$

$$\hat{Q}_3^{(1)} = \frac{1}{2}\sum_j (\vec{\sigma}_{j+1} + \vec{\sigma}_{j+2}) \cdot (\vec{\sigma}_j \times \vec{\sigma}_{j+3}). \tag{69}$$

This operator is used to define a quasi-conserved quantity $\hat{Q}_3^{(0)} + \lambda\hat{Q}_3^{(1)}$ for the perturbed Hamiltonian. In fact, it satisfies

$$[\hat{Q}_3^{(0)}, \hat{V}] + [\hat{Q}_3^{(1)}, \hat{H}_0] = 0, \tag{70}$$

and therefore, $\hat{Q}_3^{(0)} + \lambda\hat{Q}_3^{(1)}$ is a quasiconserved quantity:

$$[\hat{Q}_3^{(0)} + \lambda\hat{Q}_3^{(1)}, \hat{H}_0 + \lambda\hat{V}] = O(\lambda^2). \tag{71}$$

In the exercise session, we will find a generator $\hat{X}$ such that we can write this perturbation as $\hat{V} = i[\hat{X}, \hat{H}_0] + \hat{D}$, thus finding a systematic construction for all the quasiconserved quantities (using $\hat{Q}_\alpha^{(1)} = i[\hat{X}, \hat{Q}_\alpha^{(0)}]$).

**Exercise 2** *Check that Eq. (70) is satisfied.*

## 5.2 Quasiconservation

As we will now show, an immediate consequence of the existence of some quasi-conserved quantities $\hat{Q}^{(0)} + \lambda\hat{Q}^{(1)}$ is that these quantities are "slow" operators, that evolve on timescales longer than $\lambda^{-2}$. Let us use $\hat{M}$ as a shorthand notation for $\hat{Q}^{(0)} + \lambda\hat{Q}^{(1)}$ and assume that it can be written as the sum of operators $\hat{m}_j$ with finite support around a site $j$ in the lattice.

$$\hat{M} = \sum_j \hat{m}_j \equiv \hat{Q}^{(0)} + \lambda\hat{Q}^{(1)}. \tag{72}$$

If $\hat{M}$ is a quasiconserved quantity, then

$$[\hat{M}, \hat{H}] = \lambda^2 [\hat{Q}^{(1)}, \hat{V}], \tag{73}$$

and the operator

$$\hat{R} = \sum_j \hat{r}_j \equiv [\hat{Q}^{(1)}, \hat{V}] \tag{74}$$

can also be written as a sum of local operators, with finite norm $\|r_j\| = C$. Then it can be proven that

$$\frac{1}{N} |\, \langle \hat{M}(t) \rangle - \langle \hat{M}(0) \rangle \,| \leq \lambda^2 C t. \tag{75}$$

**Proof:**

$$\hat{M}(t) - \hat{M}(0) = \int_0^t ds\, \frac{d}{ds} \hat{M}(s) \tag{76}$$

$$= \int_0^t ds\, e^{i\hat{H}s} i[\hat{H}, \hat{M}] e^{-i\hat{H}s} \tag{77}$$

$$= -\lambda^2 \int_0^t ds\, e^{i\hat{H}s} \hat{R} e^{-i\hat{H}s} \tag{78}$$

$$= -\lambda^2 \sum_j \int_0^t ds\, e^{i\hat{H}s} \hat{r}_j e^{-i\hat{H}s}. \tag{79}$$

We now evaluate this on the initial state $|\psi_0\rangle$ and we get

$$\frac{1}{N} |\, \langle \hat{M}(t) \rangle - \langle \hat{M}(0) \rangle \,| = \frac{\lambda^2}{N} \left| \sum_j \int_0^t ds\, \langle e^{i\hat{H}s} \hat{r}_j e^{-i\hat{H}s} \rangle \right| \tag{80}$$

$$\leq \frac{\lambda^2}{N} \sum_j \left| \int_0^t ds\, \langle e^{i\hat{H}s} \hat{r}_j e^{-i\hat{H}s} \rangle \right| \leq \lambda^2 C t. \qquad \square \tag{81}$$

For the original conserved quantity $\hat{Q}^{(0)}$, instead, the dynamics generally happens on faster timescales, of order $\lambda^{-1}$, as estimated in Eq. (52).

Note that these bounds are very general: they hold for any initial state and any time $t$. In the next subsession, we will consider, instead, a more specific scenario: the one where Fermi's golden rule is applicable.

### 5.3 Vanishing rate

We now show that the Fermi's golden rule rate vanishes for weak perturbations. To do so, we take the condition Eq. (59) and compute the matrix elements on a basis of simultaneous eigenstates of $\hat{H}_0$, $\hat{Q}^{(0)}$:

$$\langle E_m^{(0)}|[\hat{Q}^{(1)},\hat{H}_0]|E_n^{(0)}\rangle = -\langle E_m^{(0)}|[\hat{Q}^{(0)},\hat{V}]|E_n^{(0)}\rangle \tag{82}$$

which gives

$$(E_m - E_n)Q_{nm}^{(1)} = (Q_{mm}^{(0)} - Q_{nn}^{(0)})V_{nm}. \tag{83}$$

Now we can use this equality to write Eq. (57) as

$$\frac{\langle\hat{Q}^{(0)}(t+\Delta t)\rangle - \langle\hat{Q}^{(0)}(t)\rangle}{\Delta t} = 2\pi\lambda^2 \sum_n \sum_{m\neq n} (Q_{nn}^{(0)} - Q_{mm}^{(0)})P_m(t)|V_{nm}|^2\delta(\omega_{nm}) \tag{84}$$

$$= -2\pi\lambda^2 \sum_n \sum_{m\neq n} (E_m - E_n)Q_{nm}^{(1)}P_m(t)V_{mn}\delta(\omega_{nm}) = 0 \tag{85}$$

It's important to note that this doesn't mean $\hat{Q}$ remains strictly constant over time. Rather, its decay isn't captured by the standard Fermi's golden rule rate, which comes from second-order perturbation theory. Instead, the decay is driven by higher-order effects that we've ignored up to this point.

#### 5.3.1 Geometric picture

The notion of weak perturbations has a simple geometric interpretation. Suppose we start from $\hat{H}_0$ and $\hat{Q}^{(0)}$ and apply a continuous family of unitaries $\hat{U}_\lambda = e^{i\lambda\hat{X}}$, with $\hat{U}_0 = \hat{I}$. We define

$$\hat{H}(\lambda) = \hat{U}_\lambda \hat{H}_0 \hat{U}_\lambda^\dagger, \quad \hat{Q}(\lambda) = \hat{U}_\lambda \hat{Q}^{(0)} \hat{U}_\lambda^\dagger. \tag{86}$$

These describe an integrable deformation of $\hat{H}_0$, sometimes called a **long-range integrable deformation**, because $\hat{H}(\lambda)$ may be long-range. The key point is that

$$[\hat{H}(\lambda), \hat{Q}(\lambda)] = 0, \tag{87}$$

so integrability is preserved under this deformation.

To first order in $\lambda$, we get:

$$\hat{H}(\lambda) = \hat{H}_0 + i\lambda[\hat{X},\hat{H}_0] + O(\lambda^2), \tag{88}$$

$$\hat{Q}(\lambda) = \hat{Q}^{(0)} + i\lambda[\hat{X},\hat{Q}^{(0)}] + O(\lambda^2). \tag{89}$$

So if we define $\hat{V} = i[\hat{X},\hat{H}_0] + \hat{D}$ (where, once again, $\hat{D}$ commutes with $\hat{H}_0$ and $\hat{Q}^{(0)}$), then

$$\hat{H}_0 + \lambda\hat{V} = \hat{H}(\lambda) + O(\lambda^2), \tag{90}$$

i.e., our perturbed system is close (up to $\lambda^2$) to an integrable one. In this sense, the breaking of integrability is parametrically weak.

This idea is illustrated in Fig. 2, which schematically illustrates the space of perturbations of an integrable model. Weak perturbations are tangent to the manifold of long-range integrable deformations.

Eq. (90) shows that the effective integrability breaking is $O(\lambda^2)$ rather than $O(\lambda)$, so we expect that the rate of thermalization scales as $\lambda^4$, rather than the usual Fermi's golden rule rate $\lambda^2$. As we will show later, this construction indeed implies that the Fermi's golden rule rate vanishes at order $\lambda^2$, leading to much slower thermalization.

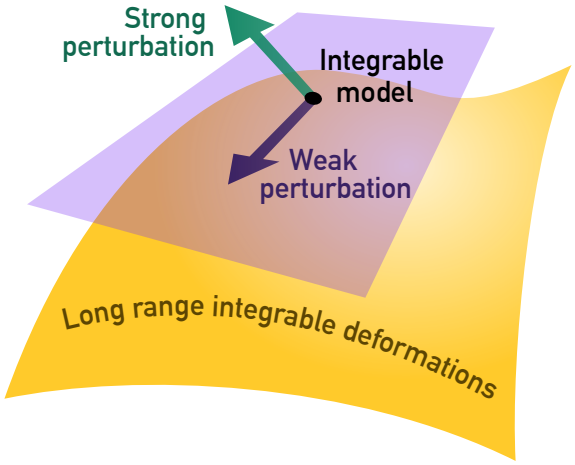


Figure 2: The yellow surface represents the manifold of long-range integrable deformations of an integrable model $\hat{H}_0$. A *weak* perturbation $\hat{V}$ is tangent to this manifold, in contrast to generic *strong* perturbations.

### 5.4 Generators

We argued above that a sufficient condition for $\hat{V}$ being a weak integrability breaking perturbation is that it can be written as $\hat{V} = i[\hat{X}, \hat{H}_0] + \hat{D}$, where $\hat{D}$ commutes with both the Hamiltonian and the conserved quantities, and the operator $\hat{X}$ can be interpreted as the generator of a unitary transformation.

This decomposition gives a nice systematic way of finding weak perturbations. All we need to do, is define the generator $\hat{X}$ (and possibly, the operator $\hat{D}$), and we have a machinery to find as many weak perturbations as we want! With one caveat: we want both $\hat{V}$ and $\hat{Q}^{(1)}$ to be extensive local. What types of generators $\hat{X}$ are guaranteed to give extensive local operators when commuted with $\hat{H}$?

#### 5.4.1 Extensive operators

The first class of generators that we will examine are extensive generators. Suppose we can write $\hat{X}$ as a sum of a local density

$$\hat{X}_{ex} = \sum_j \hat{o}_j. \tag{91}$$

Then the perturbation $\hat{V}_{ex} = i[\hat{X}_{ex}, \hat{H}_0]$ and the corresponding corrections to the charges $i[\hat{X}_{ex}, \hat{Q}^{(0)}_\alpha]$ are also sums of local terms. These operators can therefore be used as generators of weak perturbations.

#### 5.4.2 Boosted operators

Another class of generators are so-called boosted generators. Given a conserved quantity of the original model $\hat{Q}^{(0)}_\beta = \sum_j \hat{q}^{(0)}_{\beta,j}$, we can define the boost of $\hat{Q}^{(0)}_\beta$ as

$$B[\hat{Q}^{(0)}_\beta] = \sum_j j\, \hat{q}^{(0)}_{\beta,j}. \tag{92}$$

The fact that these operators can be used as generators of weak perturbations is not obvious. Why should the commutator between $\hat{X}_{bo} = B[\hat{Q}^{(0)}_\beta]$ and the conserved quantities $\hat{Q}^{(0)}_\alpha$ (including the Hamiltonian) give local extensive operators? The trick here is in the generalized currents defined above. In fact, we can write the commutator as

$$i[\hat{X}_{bo},\hat{Q}^{(0)}_\alpha] = i[B[\hat{Q}^{(0)}_\beta],\hat{Q}^{(0)}_\alpha] \tag{93}$$
$$= i\sum_j j[\hat{q}^{(0)}_{\beta,j},\hat{Q}^{(0)}_\alpha] \tag{94}$$
$$= -\sum_j j(\hat{J}_{\beta\alpha,j} - \hat{J}_{\beta\alpha,j+1}) \tag{95}$$
$$= -\sum_j [j\hat{J}_{\beta\alpha,j} - (j-1)\hat{J}_{\beta\alpha,j}] \tag{96}$$
$$= -\sum_j \hat{J}_{\beta\alpha,j}. \tag{97}$$

Therefore, even though $\hat{X}_{bo}$ is not extensive local, it gives extensive local operator when commuted with any conserved quantity (including the Hamiltonian). Therefore, it can be used to generate weak perturbations.

### 5.4.3 Bilocal operators

We will now consider an even more exotic class of generators, called bilocal operators. Given two conserved quantities $\hat{Q}_\beta$ and $\hat{Q}_\gamma$, we define the bilocal operator, denoted as $[\hat{Q}_\beta|\hat{Q}_\gamma]$:

$$[\hat{Q}_\beta|\hat{Q}_\gamma] = \sum_{j<k}\{\hat{q}_{\beta,j},\hat{q}_{\gamma,k}\} + \frac{1}{2}\sum_j \{\hat{q}_{\beta,j},\hat{q}_{\gamma,j}\}. \tag{98}$$

We aim to use it as a generator $\hat{X}_{bi} \equiv [\hat{Q}_\beta|\hat{Q}_\gamma]$ for weak integrability-breaking perturbations. We thus have to show that $i[\hat{X}_{bi},\hat{Q}_\alpha]$ is extensive and local.

**Exercise 3** *Use the definition of the generalized currents Eq. (50) to show that*

$$i[\hat{X}_{bi},\hat{Q}_\alpha] = i[[\hat{Q}_\beta|\hat{Q}_\gamma],\hat{Q}_\alpha] = \frac{1}{2}\sum_j \{\hat{q}_{\gamma,j},\hat{J}_{\beta\alpha,j}+\hat{J}_{\beta\alpha,j+1}\} - \frac{1}{2}\sum_j \{\hat{q}_{\beta,j},\hat{J}_{\gamma\alpha,j}+\hat{J}_{\gamma\alpha,j+1}\}. \tag{99}$$

*This proves that $\hat{X}_{bi}$ is a good generator.*

## 5.5 The adiabatic gauge potential

We showed so far how to construct weak integrability breaking perturbations by considering various classes of generators. We may often be interested in understanding whether a given perturbation breaks integrability weakly. Given $\hat{V}$, can we write in the form

$$\hat{V} = i[\hat{X},\hat{H}_0] + \hat{D}, \tag{100}$$

with $[\hat{D},\hat{H}_0] = 0$?

It naively seems that the answer to this question is always positive. We can write the equation in the basis of energy eigenstates of $\hat{H}_0$ (or better, a basis where both $\hat{H}_0$ and $\hat{D}$ are diagonal):

$$V_{nm} = iX_{nm}(E_m - E_n) + D_n\delta_{nm}. \tag{101}$$

We can define $D_n \equiv V_{nn}$ and

$$X_{nm} \equiv \begin{cases} \frac{V_{nm}}{i(E_m-E_n)} & \text{for } E_n \neq E_m \\ 0 & \text{otherwise,} \end{cases} \tag{102}$$

and Eq. (100) is satisfied. The operator defined by Eq. (102) was introduced in other contexts to study the onset of thermalization, and is called the adiabatic gauge potential.

The problem with this result is that the generator obtained in Eq. (102) may be very non-local, and is not guaranteed, in general, to give a local extensive correction $\hat{Q}^{(1)}$ to the quasiconserved charges. In fact, the denominator $E_m - E_n$ can be as small as $\sim \mathcal{D}_N^{-1}$ (where $\mathcal{D}_N$ is the Hilbert space dimension), while the matrix elements of a generic operator $V_{nm}$ are typically expected to be of order $\sim \mathcal{D}_N^{-1/2}$. Therefore, the norm of $\hat{X}^{AGP} \equiv \sum_{nm} X_{nm} |E_n\rangle \langle E_m|$ is typically expected to be exponentially large in the system size. This was found numerically for strong perturbations of integrable models[3] [3,4]. For weak perturbations, on the other hand, the norm of the adiabatic gauge potential $\hat{X}^{AGP}$ was shown to grow only polynomially with $N$.

# 6 Exponentially slow thermalization

So far we found that the usual theory for describing prethermalization under a Hamiltonian $\hat{H} = \hat{H}_0 + \lambda \hat{V}$ predicts that the evolution towards thermal equilibrium occurs at a rate given by Fermi's golden rule. We also showed that there are classes of perturbations that have a vanishing Fermi's golden rule rate and whose thermalization rate depends on a higher power of $\lambda$. In this section we will show some examples of models that have a provably much smaller thermalization time, exponentially small in $\lambda^{-1}$.

The underlying idea is similar to the one of weak perturbations: in that case, we found that there exists a unitary transformation that "removes" the perturbation of order $\lambda$ and introduces one at order $\lambda^2$.

$$\hat{U}(\hat{H}_0 + \lambda \hat{V})\hat{U}^\dagger = \tilde{H} + O(\lambda^2) \tag{103}$$

In the cases that will be discussed in this session, this construction can be further repeated, pushing the perturbation to higher and higher orders in $\lambda$. This construction breaks down at an order $n^* \sim \lambda^{-1}$. Therefore, the actual perturbation is of order $\exp(-1/\lambda)$.

$$\hat{U}(\hat{H}_0 + \lambda \hat{V})\hat{U}^\dagger = \tilde{H} + O(\exp(-1/\lambda)) \tag{104}$$

Here we will not show the proofs for these results, which are quite involved, and refer the interested reader to Refs. [5] and [6]. We will only give some informal statements of the theorems.

## 6.1 Hamiltonians with integer spectra

A theorem proved in [5] states that, if $\hat{H}_0$ has integer, spectrum, i.e., all its eigenvalues are integer multiples of a quantity $\Delta$ (they may be degenerate), then there exists a quasilocal unitary transformation $\hat{U}$ such that

$$\hat{U}(\hat{H}_0 + \lambda \hat{V})\hat{U}^\dagger = \tilde{H} + O(\exp(-1/\lambda)), \tag{105}$$

where $\tilde{H}$ has a block-diagonal structure and only acts within the degenerate blocks of $\hat{H}_0$.

A primary example is the Hubbard model, in arbitrary dimensions:

$$\hat{H} = J \sum_{\langle x,y\rangle,\sigma} \hat{c}^\dagger_{x,\sigma} \hat{c}_{y,\sigma} + U \sum_x \hat{n}_{x,\downarrow} \hat{n}_{x,\uparrow}. \tag{106}$$

If $U \gg J$, we can treat the onsite interaction as $\hat{H}_0$ and the hopping as the perturbation. The eigenvalues of $\hat{H}_0 = U \sum_x \hat{n}_{x,\downarrow} \hat{n}_{x,\uparrow}$ are integer multiples of $U$ and count the number of double

[3] With the exception of perturbations to free fermion models, see Ref. [2].

occupancies $\hat{N} = \sum_x \hat{n}_{x,\downarrow}\hat{n}_{x,\uparrow}$ (also called *doublons*) in the system. Then the theorem implies that $[\hat{N}, \tilde{H}] = 0$ and therefore $[\hat{U}^\dagger \hat{N} \hat{U}, \hat{H}_0 + \lambda \hat{V}] = O(\exp(-1/\lambda))$. The quantity $\hat{U}^\dagger \hat{N} \hat{U}$ is therefore a quasiconserved quantity up to times that are exponentially long in $\lambda^{-1}$.

### 6.2 False vacuum decay

A theorem in [6] applies to a Hamiltonian $\hat{H}_0$ with a gap $\Delta$ in the spectrum. The ground state does not have to be unique: more generally, the Hamiltonian $\hat{H}_0$ is split into a low-energy and a high-energy sector, separated by $\Delta$. Then there exists a quasilocal unitary transformation $\hat{U}$ such that

$$\hat{U}(\hat{H}_0 + \lambda \hat{V})\hat{U}^\dagger = \tilde{H} + O(\exp(-1/\lambda^a)), \tag{107}$$

where $\tilde{H}$ has a block-diagonal structure that preserves the block structure of the low and high energy sectors, and $a > 0$ depends on the dimensionality.

An application of the theorem is in the context of the decay of a false vacuum. Let us consider, for example, the quantum Ising chain

$$\hat{H}_0 = \sum_j (-J\sigma_j^z \sigma_{j+1}^z + h\sigma_j^x). \tag{108}$$

In the ferromagnetic phase, for $J > |h|$, the model is gapped, with two degenerate ground states $|\psi_\uparrow\rangle, |\psi_\downarrow\rangle$ with opposite magnetization $\langle\psi_\uparrow|\hat{\sigma}_j^z|\psi_\uparrow\rangle = -\langle\psi_\downarrow|\hat{\sigma}_j^z|\psi_\downarrow\rangle = M$. we now perturb the systems with a longitudinal field

$$\hat{V} = \sum_j \sigma_j^z. \tag{109}$$

This perturbation explicitly breaks the $Z_2$ symmetry of the Ising chain $\hat{H}_0$ and splits the degeneracy of the two ground states in a dramatic way: the energy difference between them becomes $\sim 2|\lambda|NM$, i.e., it is extensive. The ground state of $\hat{H} = \hat{H}_0 + \lambda\hat{V}$ is non-degenerate and corresponds to one of the original ground states of $\hat{H}_0$, while the other state becomes a highly excited state, and, as shown in [6], is long-lived with a lifetime that is exponentially long in $\lambda^{-1}$. This state is often called a metastable state or a false vacuum.

## Acknowledgements

## A Fermi's Golden rule

## B Lieb-Robinson bound

## C Bound on time evolved observables

$$|\langle\psi_0| e^{iHt} O e^{-iHt} |\psi_0\rangle - \langle\psi_0| e^{iH_0 t} O e^{-iH_0 t} |\psi_0\rangle| \tag{C.1}$$
$$= |\langle\psi_0| e^{iH_0 t}(O_{\rm int}(t) - O) e^{-iH_0 t} |\psi_0\rangle| \tag{C.2}$$
$$\leq \|O_{\rm int}(t) - O\|, \tag{C.3}$$

where we defined the time evolved operator in the interaction picture

$$O_{\rm int}(t) = e^{-iH_0 t} e^{iHt} O e^{-iHt} e^{iH_0 t}, \tag{C.4}$$

whose evolution is determined by the equation

$$\frac{d}{dt}O_{\rm int}(t) = i\lambda[V(t), O_{\rm int}(t)] \tag{C.5}$$

$$\|O_{\rm int}(t) - O\| = \lambda \| \int_0^t ds [V(s), O_{\rm int}(s)]\| \leq \lambda \int_0^t ds \|[V(s), O_{\rm int}(s)]\| = \lambda \int_0^t ds \|[V, e^{iHt} O e^{-iHt}]\| \tag{C.6}$$

$V(s) = e^{-iH_0 t} V e^{iH_0 t}$.

Now we can use the Lieb-Robinson bound:

$$\|[A, e^{itH} B e^{-itH}]\| \leq \|A\| \|B\| e^{-\kappa(d(X,Y)-vt)} \min(|X|, |Y|). \tag{C.7}$$

So we can use this bound out of the light cone, and the trivial bound in the light cone, to prove that that quantity is bounded by $t\times$ constant.

# D Another derivation of the kinetic equation

Assumption

$$\rho(t) = \rho_{GGE}(t) + \delta\rho(t) \tag{D.1}$$

Interaction picture

$$\partial_t \rho_I(t) = -i\lambda[V_I(t), \rho_I(t)] \tag{D.2}$$

$$\frac{d\langle Q_\alpha^{(0)}\rangle}{dt} = -i\lambda {\rm Tr}[Q_\alpha^{(0)}[V_I(t), \rho_I(t)]] = -i\lambda {\rm Tr}[[Q_\alpha^{(0)}, V_I(t)]\rho_I(t)] = -i\lambda \langle [Q_\alpha^{(0)}, V_I(t)]\rangle \tag{D.3}$$

If we approximate

$$\frac{d\langle Q_\alpha^{(0)}\rangle}{dt} \approx -i\lambda {\rm Tr}[[Q_\alpha^{(0)}, V_I(t)]\rho_{GGE}(t)] = 0 \tag{D.4}$$

We need to include the next order correction

$$\rho_I(t) = \rho(0) - i\lambda \int_0^t ds [V_I(s), \rho_I(s)] \approx \rho(0) - i\lambda \int_0^t ds [V_I(s), \rho_{GGE}(s)] \tag{D.5}$$

Then

$$\frac{d\langle Q_\alpha^{(0)}\rangle}{dt} \approx -i\lambda {\rm Tr}[[Q_\alpha^{(0)}, V_I(t)]\rho(0)] - \lambda^2 \int_0^t ds\, {\rm Tr}[[Q_\alpha^{(0)}, V_I(t)], [V_I(s), \rho_{GGE}(s)]] \tag{D.6}$$

$$\frac{d\langle Q_\alpha^{(0)}\rangle}{dt} \approx -\lambda^2 \int_0^t ds\, \langle [[Q_\alpha^{(0)}, V_I(t)], V_I(s)]\rangle_{GGE,s} \tag{D.7}$$

$$\frac{d\langle Q_\alpha^{(0)}\rangle}{dt} \approx -2\pi\lambda^2 \sum_{n,m} (\rho_{GGE})_{nn} [(Q_\alpha^{(0)})_{nn} - (Q_\alpha^{(0)})_{mm}] |V_{nm}|^2 \frac{\sin(E_n - E_m)\Delta t}{\pi(E_n - E_m)} \tag{D.8}$$

$$\frac{d\langle Q_\alpha^{(0)}\rangle}{dt} \approx -2\pi\lambda^2 \sum_{n,m} (\rho_{GGE})_{nn} [(Q_\alpha^{(0)})_{nn} - (Q_\alpha^{(0)})_{mm}] |V_{nm}|^2 \delta(E_n - E_m) \tag{D.9}$$

$$\frac{d\langle Q_\alpha^{(0)}\rangle}{dt} \approx -2\pi\lambda^2 \sum_{n,m} (Q_\alpha^{(0)})_{nn} [(\rho_{GGE})_{nn} - (\rho_{GGE})_{mm}] |V_{nm}|^2 \delta(E_n - E_m) \tag{D.10}$$

## References


[1] K. Mallayya, M. Rigol and W. De Roeck, *Prethermalization and thermalization in isolated quantum systems*, Phys. Rev. X **9**, 021027 (2019), doi:10.1103/PhysRevX.9.021027.

[2] B. Pozsgay, R. Sharipov, A. Tiutiakina and I. Vona, *Adiabatic gauge potential and integrability breaking with free fermions*, SciPost Phys. **17**, 075 (2024), doi:10.21468/SciPostPhys.17.3.075.

[3] P. Orlov, A. Tiutiakina, R. Sharipov, E. Petrova, V. Gritsev and D. V. Kurlov, *Adiabatic eigenstate deformations and weak integrability breaking of heisenberg chain*, Phys. Rev. B **107**, 184312 (2023), doi:10.1103/PhysRevB.107.184312.

[4] S. Vanovac, F. M. Surace and O. I. Motrunich, *Finite-size generators for weak integrability breaking perturbations in the heisenberg chain*, Phys. Rev. B **110**, 144309 (2024), doi:10.1103/PhysRevB.110.144309.

[5] D. Abanin, W. De Roeck, W. W. Ho and F. Huveneers, *A rigorous theory of many-body prethermalization for periodically driven and closed quantum systems*, Communications in Mathematical Physics **354**(3), 809 (2017), doi:https://doi.org/10.1007/s00220-017-2930-x.

[6] C. Yin and A. Lucas, *Prethermalization and the local robustness of gapped systems*, Phys. Rev. Lett. **131**, 050402 (2023), doi:10.1103/PhysRevLett.131.050402.

## Exercise session 3

### 1. Extensive generator

Let us consider a simple extensive generator and use it to construct a weak perturbation of the Heisenberg model. We define

$$\hat{X}_{\text{ex}} = \sum_j \vec{\sigma}_j \cdot \vec{\sigma}_{j+2}. \tag{D.11}$$

- What are the properties of $\hat{X}_{ex}$ under inversion symmetry and time-reversal symmetry?
- Compute $\hat{V}_{\text{ex}} = i[\hat{X}_{\text{ex}}, \hat{H}_0]$.
- What are the properties of $\hat{V}_{ex}$ under inversion and under time-reversal symmetry?
- Find an extensive generator $\hat{X}'_{ex}$ that can be used to construct a perturbation $\hat{V}'_{ex} = i[\hat{X}'_{ex}, \hat{H}_0]$ that is invariant under inversion and time reversal (no need to compute $V'_{ex}$).

### 2. Boosted generator

Now we consider a boosted generator

$$\hat{X}_{bo} = \sum_j j\, q^{(0)}_{j,3} = -\frac{1}{2}\sum_j j\,(\vec{\sigma}_j \times \vec{\sigma}_{j+1}) \cdot \vec{\sigma}_{j+2}. \tag{D.12}$$

- Compute the generalized current $\hat{J}_{32,j}$. Hint: use that $i[\vec{\sigma}_j \cdot \vec{\sigma}_k, (\vec{\sigma}_j \times \vec{\sigma}_k)^a] = 4(\vec{\sigma}_j - \vec{\sigma}_k)^a$
- Compute $\hat{V}_{bo} = i[\hat{X}_{bo}, \hat{H}_0]$ (use Eq. (93)).
- What are the properties of $\hat{V}_{bo}$ under inversion and time-reversal?
- Find an operator $\hat{D}$, with $[\hat{D}, \hat{H}_0] = 0$, such that $\hat{V}_{bo} + \hat{D} = -\sum_j \vec{\sigma}_j \cdot \vec{\sigma}_{j+2}$.