Einstein's kinetic theory of the Brownian motion, based upon light water molecules continuously bombarding a heavy pollen, provided an explanation of diffusion from the Newtonian mechanics. Since the discovery of quantum mechanics it has been a challenge to verify the emergence of diffusion from the Schrödinger equation.

The first step in this program is to verify the linear Boltzmann equation as a certain scaling limit of a Schrödinger equation with random potential. In the second step, one considers a longer time scale that corresponds to infinitely many Boltzmann collisions. The intuition is that the Boltzmann equation then converges to a diffusive equation similarly to the central limit theorem for Markov processes with sufficient mixing. In these lecture notes (prepared for the Les Houches summer school in 2010 August) we present the mathematical tools to rigorously justify this intuition. The new material relies on joint papers with H.-T. Yau and M. Salmhofer.

AMS Subject Classification: 60J65, 81T18, 82C10, 82C44

Contents

1 Overview of the rigorous derivations of diffusions

The fundamental equations governing the basic laws of physics, the Newton and the Schrödinger equations, are time reversible and have no dissipation. It is remarkable that dissipation is nevertheless ubiquitous in nature, so that almost all macroscopic phenomenological equations are dissipative. The oldest such example is perhaps the equation of heat conductance found by Fourier. This investigation has led to the heat equation, the simplest type of diffusion equations:

$$\partial_t u = \Delta_x u, \tag{1.1}$$

where $u(x,t)$ denotes the temperature at position $x \in \mathbb{R}^d$ and time t. One key feature of the diffusion equations is their inhomogeneous scaling; the time scales as the square of the spatial distance:

$$t \sim x^2; \qquad \text{time} \sim (\text{distance})^2.$$

In these lectures we will explore how diffusion equations emerge from first principle physical theories such as the classical Newtonian dynamics and the quantum Schrödinger dynamics. In Section 1 we give an overview of existing mathematical results on the derivation of diffusion (and related) equations. In Sections 2–4 we discuss the basic formalism and present a few well-known preliminary facts on stochastic, classical and quantum dynamics. An experienced reader can skip these sections. In Section 5 we introduce our main model, the random Schrödinger equation or the quantum Lorentz gas and its lattice version, the Anderson model. In Section 6 we formulate our main theorems that state that the random Schrödinger equation exhibits diffusion and after a certain scaling and limiting procedure it can be described by a heat equation.

The remaining sections contain the sketch of the proofs. Since these proofs are quite long and complicated, we will not only have to omit many technical details, but even several essential ideas can only be mentioned very shortly. We will primarily focus on the most important aspect: the classification and the estimates of the so-called non-repetition Feynman graphs. Estimates of other Feynman graphs, such as recollision graphs and graphs with higher order pairings will only be discussed very superficially.

Our new results presented in this review were obtained in collaboration with H.-T. Yau and M. Salmhofer.

Acknowledgement. The author is grateful to M. Salmhofer for many suggestions to improve this presentation and for making his own lecture notes available from which, in particular, the material of Section 7.2 and 8.3 was borrowed.

1.1 The founding fathers: Brown, Einstein, Boltzmann

The story of diffusion starts with R. Brown in 1827 who observed almost two centuries ago that the motion of a wildflower pollen suspended in water was erratic [6]. He saw a picture similar

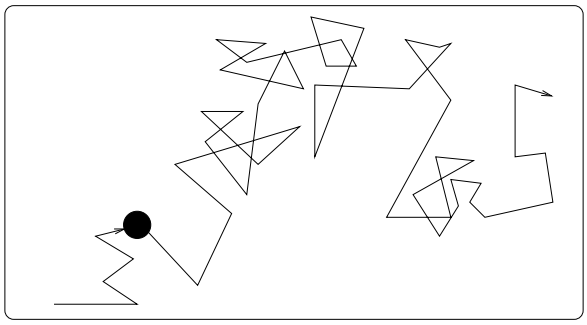

Figure 1: Brown's picture under the microscope

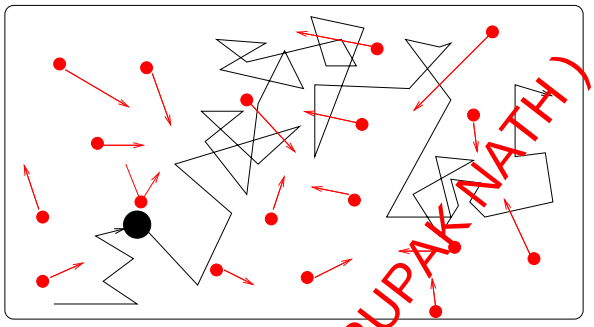

Figure 2: Einstein's explanation to Brown's picture

to Fig. 1 under his microscope (the abrupt velocity changes in this picture are exaggerated to highlight the effect, see later for more precise explanation). He "observed many of them very evidently in motion" and argued that this motion was not due to water currents but "belonged to the particle itself". First he thought that this activity characterizes only "living" objects like pollens, but later he found that small rock or glass particles follow the same pattern. He also noted that this apparently chaotic motion never seems to stop.

This picture led to the kinetic explanation by A. Einstein in 1905 [15] that such a motion was created by the constant kicks on the relatively heavy pollen by the light water molecules. Independently, M. Smoluchowski [44] arrived at a similar theory. It should be noted that at that time even the atomic-molecular structure of matter was not universally accepted. Einstein's theory was strongly supported by Boltzmann's kinetic theory, which, however, was phenomenological and seriously debated at the time. Finally in 1908 Perrin [41] (awarded the Nobel prize in 1926) experimentally verified Einstein's theory and used it, among others, to give a precise estimate on the Avogadro number. These experiments gave the strongest evidence for atoms and molecules at that time.

We should make a remark on Boltzmann's theory although this will not be the subject of this lecture. Boltzmann's kinetic theory postulates that in a gas of interacting particles at relatively low density the collisions between the particles are statistically independent (Ansatz of *molecular chaos*). This enabled him to write up the celebrated (nonlinear) Boltzmann equation for the time evolution of the single particle phase space density $f_t(x, v)$. The main

assumption is that the collision rate is simply the product of the particle densities with the appropriate velocities, thus he arrived at the following equation:

$$\partial_t f_t(x,v) + v \cdot \nabla_x f_t(x,v) = \int \sigma(v,v_1)\big[f_t(x,v')f_t(x,v_1') - f_t(x,v)f_t(x,v_1)\big]. \tag{1.2}$$

Here v, v_1 is the pair of incoming velocities of the colliding particles and v', v_1' are the outgoing velocities. In case of elastic hard balls, the pair (v', v_1') is uniquely determined by (v, v_1) (plus a randomly chosen contact vector) due to energy and momentum conservation, for other type of collisions the kernel $\sigma(v, v_1)$ describes microscopic details of the collision mechanism. The integration is for the random contact vector and all other variables apart from x, v, subject to constraints (momentum and energy conservation) encoded in $\sigma(v, v_1)$.

In addition to the highly non-trivial Ansatz of independence, this theory also assumes the molecular nature of gases and fluids in contrast to their more intuitive continuous description. Nowadays we can easily accept that gases and fluids on large scales (starting from micrometer scales and above – these are called *macroscopic scales*) can be characterized by continuous density functions, while on much smaller scales (nanometers and below, these are called *microscopic scales*) the particle nature dominates. But at Boltzmann's time the particle picture was strongly debated. The ingenuity of Boltzmann's theory is that he could combine these two pictures. The Boltzmann equation is an equation of continuous nature (as it operates with density functions) but its justification (especially the determination of the collision kernel) follows an argument using particles. No wonder that it gave rise to so much controversy especially before experiments were available. Moreover, Boltzmann theory implies irreversibility (entropy production) that was for long thought to be more a philosophical than a physical question.

After this detour on Boltzmann we return to the diffusion models. Before we discuss individual models, let me mention the *key conceptual difficulty* behind all rigorous derivations of diffusive equations. Note that the Hamiltonian dynamics (either classical or quantum) is reversible and deterministic. The diffusion equation (1.1), and also the Boltzmann equation, is irreversible: there is permanent loss of information along their evolution (recall that the heat equation is usually not solvable backward in time). Moreover, these equations exhibit a certain chaotic nature (Brown's picture). How can these two contradicting facts be reconciled?

The short answer is that the loss of information is due to a *scale separation* and the *integration of microscopic degrees of freedom*. On the microscopic (particle) level the dynamics remains reversible. The continuous (fluid) equations live on the macroscopic (fluid) scale: they are obtained by neglecting (more precisely, integrating out) many degrees of freedom on short scales. Once we switch to the fluid description, the information in these degrees of freedom is lost forever.

This two-level explanation foreshadows that for a rigorous mathematical description one would need a scale separation parameter, usually called ε that describes the ratio between the typical microscopic and macroscopic scales. In practice this ratio is of order

$$\varepsilon = \frac{1 \ \text{Angstrom}}{1 \ \text{cm}} = 10^{-8} \ ,$$

but mathematically we will consider the $\varepsilon \to 0$ so-called *scaling limit*.

Once the scale separation is put on a solid mathematical ground, we should note another key property of the macroscopic evolution equations we will derive, namely their Markovian character. Both the heat equation (1.1) and the nonlinear Boltzmann equation (1.2) give the time derivative of the macroscopic density at any fixed time t in terms of the density at the *same time* only. I.e. these evolution equations express a process where the future state depends only on the present state; the dependence of the future on the past is only indirect through the present state. We will call this feature *Markovian property* or in short *Markovity*. Note that colliding particles do build up a memory along their evolution: the state of the recolliding particles will remember their previous collisions. This effect will have to be supressed to obtain a Markovian evolution equation, as it was already recognized by Boltzmann in his Ansatz. There are essentially two ways to reduce the rate of recollisions: either by going to a low density situation or by introducing a small coupling constant.

Apart from the rigorous formulation of the scaling limit, we thus will have to cope with the main technical difficulty: **controlling memory (recollision) effects.** Furthermore, specifically in quantum models, we will have to **control interference effects** as well.

Finally, we should remark that Einstein's model is simpler than Boltzmann's, as light particles do not interact (roughly speaking, Boltzmann's model is similar to Fig. 2 but the light particles can also collide with each other and not only with the heavy particle). Thus the verification of the Ansatz of molecular chaos from first principle Hamiltonian dynamics is technically easier in Einstein's model. Still, Einstein's model already addresses the key issue: how does diffusion emerge from Hamiltonian mechanics?

1.2 Mathematical models of diffusion

The first step to setup a correct mathematical model is to recognize that some stochasticity has to be added. Although we expect that for a **"typical"** initial configuration the diffusion equations are correct, this certainly will **not hold for all** initial configuration. We expect that after opening the door between a warm and a cold room, the temperature will equilibrate (thermalize) but certainly there exists a "bad" initial configuration of the participating $N \sim 10^{23}$ particles such that all "warm" particles will, maybe after some time, head towards the cold room and vice versa, i.e. the two room temperatures will be exchanged instead of thermalization. Such configuration are extremely rare; their measure in all reasonable sense goes to zero very fast as $N \to \infty$, but nevertheless they are there and prevent us from deriving diffusion equations for all initial configurations. It is, however, practically hopeless to describe all "bad" initial configurations; they may not be recognizable by glancing at the initial state. The stochastic approach circumvents this problem by precisely formulating what we mean by saying that the "bad events are rare" without identifying them.

The stochasticity can be added in several different ways and this largely determines the technical complications involved in the derivation of the diffusion. In general, more stochasticity is added, the easier the derivation is.

The easiest is if the **dynamics itself is stochastic**; the typical example being the classical random walk and its scaling limit, the Wiener process (theorized by Wiener [47] in 1923). In the typical examples, the random process governing the dynamics of the system has no correlation. The whole microscopic dynamics is Markovian (in some generalizations the dynamics

of the one particle densities is not fully Markovian, although it typically has strong statistical mixing properties). Of course this dynamics is not Hamiltonian (no energy conservation and no reversibility), and thus it is not particularly surprising that after some scaling limit one obtains a diffusion equation. The quantum counterpart of the classical random walk is a Schrödinger equation with a Markovian time dependent random potential, see [42] and the recent proof of quantum diffusion in this model by Kang and Schenker [32].

The next level of difficulty is a Hamiltonian system with a **random Hamiltonian**. Some physical data in the Hamiltonian (e.g. potential or magnetic field) is randomly selected (describing a disordered system), but then they are frozen forever and then the Hamiltonian dynamics starts (this is in sharp contrast to the stochastic dynamics, where the system is subject to fresh random inputs along its evolution). A typical model is the *Lorentz gas*, where a single Hamiltonian particle is moving among fixed random scatterers. Recollisions with the same scatterer are possible and this effect potentially ruins the Markovity. Thus this model is usually considered in the weak coupling limit, i.e. the coupling parameter λ between the test-particle and the random scatterers is set to converge to zero, $\lambda \to 0$, and simultaneously a long time limit is considered. For a weak coupling, the recollision becomes a higher order effect and may be neglected. It turns out that if

$$t \sim \lambda^{-2} \tag{1.3}$$

then one can prove a nontrivial Markovian diffusive limit dynamics (see Kesten and Papanicolaou [33], extended more recently to a bit longer times by Komorowski and Ryzhik [35]). The relation (1.3) between the coupling and the time scale is called the *van Hove limit*. Similar supressing of the recollisions can be achieved by the *low density limit* (see later).

The corresponding quantum model (*Anderson model*) is the main focus of the current lecture notes. We will show that in the van Hove limit the linear Boltzmann equation arises after an appropriate rescaling and for longer time scales the heat equation emerges. In particular, we describe the behavior of the quantum evolution in the presumed *extended states regime* of the Anderson model up to a certain time scale. We will use a mathematically rigorous perturbative approach. We mention that supersymmetric methods [16] offer a very attractive alternative approach to study quantum diffusion, although the mathematical control of the resulting functional integrals is difficult. Effective models that are reminiscent to the saddle point approximations of the supersymmetric approach are more accessible to rigorous mathematics. Recently Disertori, Spencer and Zirnbauer have proved a diffusive estimate for the two-point correlation functions in a three dimensional supersymmetric hyperbolic sigma model [13] at low temperatures and localization was also established in the same model at high temperatures [12].

The following level of difficulty is a **deterministic Hamiltonian with random initial data of many degrees of freedom**. The typical example is Einstein's model, where the test-particle is immersed in a heat bath. The heat bath is called *ideal gas*, if it is characterized by a Hamiltonian H_{bath} of non-interacting particles. The initial state is typically a temperature equilibrium state, $\exp(-\beta H_{bath})$ i.e. the initial data of the heat-bath particles are given by this equilibrium measure at inverse temperature β. Again, some scaling limit is necessary to reduce the chance of recollisions, one can for example consider the limit $m/M \to \infty$, where M

and m are the mass of the test-particle and the heat-bath particles, respectively. In this case, a collision with a single light particle does not have sizable effect on the motion of the heavy particle (this is why the abrupt changes in velocity in Fig. 1 are exaggeration). Therefore, the rate of collisions has to be increased in parallel with $m/M \to 0$ to have a sizeable total collision effect. Similarly to the van Hove limit, there is a natural scaling limit, where nontrivial limiting dynamics was proven by Dürr, Goldstein and Lebowitz [14]. On the quantum side, this model corresponds to a quantum particle immersed in a phonon (or possibly photon) bath. On the kinetic scale the Boltzmann equation was proven in [18]. Recently De Roeck and Fröhlich [10] (see also [11] for an earlier toy model) have proved quantum diffusion for a related model in $d \geq 4$ dimensions, where the mass of the quantum particle was large and an additional internal degree of freedom ("spin") was coupled to the heat bath to enhance the decay of time correlations.

We remark that the problem becomes enormously more complicated if the **heat-bath particles can interact among each other**, i.e. if we are facing a truly interacting many-body dynamics. In this case there is even no need to distinguish between the tracer particle and the heat-bath particles, in the simplest model one just considers identical particles interacting via a two-body interaction and one investigates the single-particle density function $f_t(x, v)$. In a certain scaling limit regime, the model should lead to the nonlinear Boltzmann equation (1.2). This has only been proven in the classical model for short time by Lanford [36]. His work appeared in 1975 and since then nobody could extend the proof to include longer time scales. We mention that the complications are partly due to the fact that the nonlinear Boltzmann equation itself does not have a satisfactory existence theory for long times. The corresponding quantum model is an unsolved problem; although via an order by order expansion [5] there is no doubt on the validity of the nonlinear Boltzmann equation starting from a weakly coupled interacting many-body model, the expansion cannot be controlled up to date. Lukkarinen and Spohn [39] have studied a weakly nonlinear cubic Schrödinger equation with a random initial data (drawn from a phonon bath) near equilibrium. They proved that the space-time covariance of the evolved random wave function satisfies a nonlinear Boltzmann-type equation in the kinetic limit, for short times. While this is still a one-particle model and only fluctuations around equilibrium is studied, its nonlinear character could be interpreted as a first step towards understanding the truly many-body Boltzmann equation.

Finally, the most challenging (classical) model is the **deterministic Hamiltonian with a random initial data of a few degrees of freedom**, the typical example being the various mathematical billiards. The simplest billiard is the hard-core periodic Lorentz gas (also called Sinai's billiard [8]), where the scatterers are arranged in a periodic lattice and a single point particle ("billiard ball") moves among these scatterers according to the usual rules of specular reflections. The methods of proofs here originate more in dynamical system than in statistical mechanics and they have a strong geometric nature (e.g. convexity of the scatterers is heavily used).

All these models have natural quantum mechanical analogues; for the Hamiltonian systems they simply follow from standard quantization of the classical Hamiltonian. These models are summarized in the table below and they are also illustrated in Fig. 3. We note that the quantum analogue of the periodic Lorentz gas is ballistic due to the Bloch waves (a certain

diagonalization procedure similar to the Fourier transform), thus in this case the classical and quantum models behave differently; the quantum case being relatively trivial.

	CLASSICAL MECHANICS	QUANTUM MECHANICS
Stochastic dynamics (no memory)	Random walk (Wiener)	Random kick model with zero time corr. potential (Pillet, Schenker-Kang)
Hamiltonian particle in a random environment (one body)	Lorentz gas: particle in random scatterers (Kesten-Papanicolaou) (Komorowski-Ryzhik)	Anderson model or quantum Lorentz gas (Spohn, Erdős-Yau, Erdős-Salmhofer-Yau Disertori-Spencer-Zirnbauer)
Hamiltonian particle in a heat bath (randomness in the many-body data)	Einstein's kinetic model (Dürr-Goldstein-Lebowitz)	Electron in phonon or photon bath (Erdős, Erdős-Adami, Spohn-Lukkarinen De Roeck–Fröhlich)
Periodic Lorentz gas (randomness in the one-body initial data)	Sinai billiard (Bunimovich-Sinai)	Ballistic (Bloch waves, easy)
Many-body interacting Hamiltonian	Nonlinear Boltzmann eq (short time: Lanford)	Quantum NL Boltzmann (unsolved)

2 Some facts on stochastic processes

Our goal is to understand the stochastic features of quantum dynamics. A natural way to do it is to compare quantum dynamics with a much simpler stochastic dynamics whose properties are well-known or much easier to establish. In this section the most important ingredients of elementary stochastic processes are collected and we will introduce the Wiener process. Basic knowledge of probability theory (random variables, expectation, variance, independence, Gaussian normal random variable, characteristic function) is assumed.

2.1 The central limit theorem

Let v_i, $i = 1, 2, \ldots$, be a sequence of independent, identically distributed (denoted by i.i.d. for brevity) random variables with zero expectation, $\mathbf{E}\, v_i = 0$ and finite variance, $\mathbf{E}\, v_i^2 = \sigma^2$. The values of v_i are either in $\mathbb{R}^d$ or $\mathbb{Z}^d$. In case of $d > 1$, the variance σ^2 is actually a matrix (called *covariance matrix*), defined as $\sigma_{ab}^2 = \mathbf{E}\, v^{(a)} v^{(b)}$ where $v = (v^{(1)}, v^{(2)}, \ldots, v^{(d)})$. The v_i's should be thought of as velocities or steps of a walking particle at discrete time i. For simplicity we

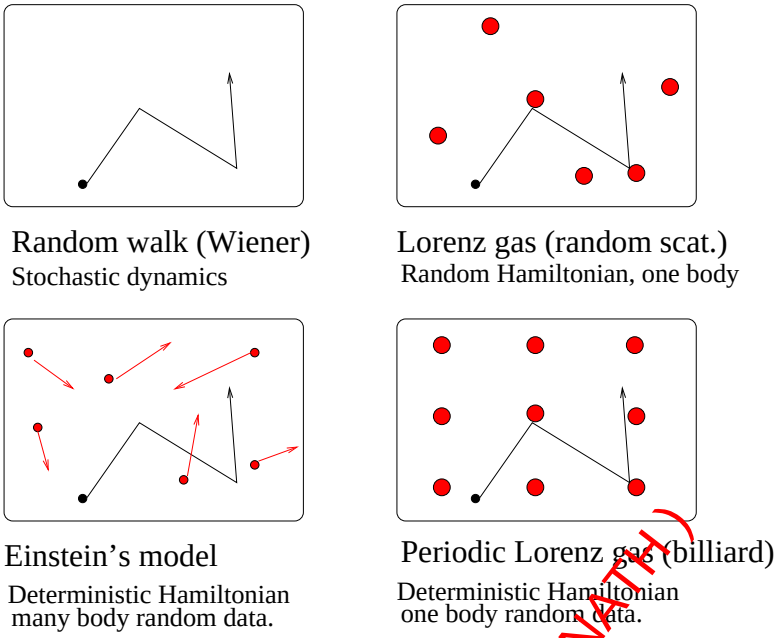

Figure 3: Microscopic models for diffusion

will work in $d = 1$ dimensions, the extension to the general case is easy. A natural question: Where is the particle after n steps if n is big?

Let

$$S_n := \sum_{i=1}^{n} v_i$$

be the location of the particle after n steps. It is clear that the expectation is zero, $\mathbf{E}\, S_n = 0$, and the variance is

$$\mathbf{E}\, S_n^2 = \mathbf{E}\Big(\sum_{i=1}^{n} v_i\Big)^2 = \sum_{i=1}^{n} \mathbf{E}\, v_i^2 = n\sigma^2 \tag{2.1}$$

(note that independence via $\mathbf{E}\, v_i v_j = 0$ for $i \neq j$ has been used).

This suggests to rescale S_n and define

$$X_n := \frac{S_n}{\sqrt{n}} = \frac{\sum_{i=1}^{n} v_i}{\sqrt{n}}$$

then $\mathbf{E}\, X_n = 0$ and $\mathbf{E}\, X_n^2 = \sigma^2$. The surprising fact is that the distribution of X_n tends to a universal one (namely to the Gaussian normal distribution) as $n \to \infty$, no matter what the initial distribution of v_i was! This is the central limit theorem, a cornerstone of probability theory and one of the fundamental theorems of Nature.

Theorem 2.1. *Under the conditions* $\mathbf{E}\, v_i = 0$ *and* $\mathbf{E}\, v_i^2 = \sigma^2 < \infty$ *the distribution of the rescaled sum* X_n *of the i.i.d random variables* v_i *converges to the normal distribution, i.e*

$$X_n \Longrightarrow X$$

in the sense of probability distributions where X is a random variable distributed according to $N(0,\sigma^2)$.

We recall that the density of the normal distribution in $\mathbb{R}^1$ is given by

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}},$$

and the convergence is the standard weak convergence of probability measures (or convergence in distribution or convergence in law):

Definition 2.2. *A sequence of random variables X_n converges to the random variable X in distribution, $X_n \Longrightarrow X$ if*

$$\mathbf{E}\, G(X_n) \to \mathbf{E}\, G(X)$$

for any continuous bounded function $G : \mathbb{R} \to \mathbb{R}$

In particular, X_n converges in distribution to $N(0,\sigma^2)$ if

$$\mathbf{E}\, G(X_n) \to \int G(x) f(x) \mathrm{d}x.$$

Analogous definitions hold in higher dimensions.

Recall that we will have two scales: a microscopic and a macroscopic one. In this example, the microscopic scale corresponds to one step of the walking particle and we run the process up to a total of n units of the microscopic time. The macroscopic time will be the natural time scale of the limit process and it will be kept order one even as $n \to \infty$. Recalling that we introduced $\varepsilon \ll 1$ as a scaling parameter, we now reformulate the central limit theorem in this language.

Let $T > 0$ be a macroscopic time (think of it as an unscaled, order one quantity) and let

$$n := [T\varepsilon^{-1}],$$

where $[\cdot]$ denotes the integer part. Let again $\mathbf{E}\, v_i = 0$, $\mathbf{E}\, v_i^2 = \sigma^2$. From the central limit theorem it directly follows that

$$\widetilde{X}_T^\varepsilon := \frac{1}{\varepsilon^{-1/2}} \sum_{i=1}^{[T\varepsilon^{-1}]} v_i \Longrightarrow X_T, \qquad \text{as } \varepsilon \to 0, \tag{2.2}$$

where X_T is a normal Gaussian variable with variance $T\sigma^2$ and with density function

$$f_T(X) = \frac{1}{\sqrt{2\pi T}\sigma} \exp\Big(-\frac{X^2}{2\sigma^2 T}\Big). \tag{2.3}$$

Note that the normalization in (2.2) is only by $\sqrt{\varepsilon^{-1}}$ and not by $\sqrt{n} \approx \sqrt{T\varepsilon^{-1}}$, therefore the limit depends on T (but not on ε, of course).

Since the Gaussian function $f_T(X)$ gives the probability of finding X_T at location x, we have

$$f_T(X)\mathrm{d}X \approx \mathrm{Prob}\{\widetilde{X}_T^\varepsilon \text{ is located at } X + \mathrm{d}X \text{ at time } T\}.$$

Note that we used macroscopic space and time coordinates, X and T. Translating this statement into the original process in microscopic coordinates we see that

$$f_T(X)\mathrm{d}X \approx \mathrm{Prob}\{\text{finding } S_n \text{ at } \varepsilon^{-1/2}(X+\mathrm{d}X) \text{ at time } n \approx \varepsilon^{-1}T\}.$$

Note that the space and time are rescaled differently; if (X,T) denote the macroscopic space/time coordinates and (x,t) denote the microscopic coordinates, then

$$x = \varepsilon^{-1/2}X, \qquad t = \varepsilon^{-1}T.$$

This is the typical *diffusive scaling*, where

$$\text{time} = (\text{distance})^2. \tag{2.4}$$

Finally, we point out that the limiting density function $f_T(X)$ satisfies the heat equation with *diffusion coefficient* σ^2:

$$\partial_T f_T(X) = \sigma^2 \partial_X^2 f_T(X). \tag{2.5}$$

In fact, the Gaussian function (2.3) is the fundamental solution to the heat equation also in higher dimensions.

The emergence of the normal distribution and the heat equation is universal; note that apart from the variance (and the zero expectation) no other information on the distribution of the microscopic step v was used. The details of the microscopic dynamics are wiped out, and only the essential feature, the heat equation (or the normal distribution) with a certain diffusion coefficient, remains visible after the $\varepsilon \to 0$ scaling limit.

We remark that the central limit theorem (with essentially the same proof) is valid if v_i's have some short time correlation. Assume that v_i is a stationary sequence of random variables with zero mean, $\mathbf{E}\, v_i = 0$, and let

$$R(i,j) := \mathbf{E}\, v_i v_j$$

the correlation function. By stationarity, $R(i,j)$ depends only on the difference, $i-j$, so let us use the notation

$$R(i-j) := \mathbf{E}\, v_i v_j$$

This function is also called the *velocity-velocity autocorrelation function.*

Theorem 2.3. *Assume that the velocity-velocity autocorrelation function is summable, i.e.*

$$\sum_{k=-\infty}^{\infty} R(k) < \infty.$$

Then the rescaled sums

$$\widetilde{X}_T^{\varepsilon} := \varepsilon^{1/2} \sum_{i=1}^{[T\varepsilon^{-1}]} v_i$$

converge in distribution to a normal random variable,

$$\widetilde{X}_T^{\varepsilon} \Longrightarrow X_T, \qquad \varepsilon \to 0,$$

whose density function $f_T(X)$ satisfies the heat equation

$$\partial_T f = D\Delta f_T$$

with a diffusion coefficient

$$D := \sum_{k=-\infty}^{\infty} R(k),$$

assuming that D is finite. This relation between the diffusion coefficient and the sum (or integral) of the velocity-velocity autocorrelation function is called the Green-Kubo formula.

Finally we mention that all results remain valid in higher dimensions, $d > 1$. The role of the diffusion coefficient is taken over by a diffusion matrix D_{ij} defined by the correlation matrix of the single step distribution:

$$D_{ij} = \mathbf{E}\, v^{(i)} v^{(j)}$$

where $v = (v^{(1)}, v^{(2)}, \ldots, v^{(d)})$. The limiting heat equation (2.5) in the uncorrelated case is modified to

$$\partial_T f_T(X) = \sum_{i,j=1}^{d} D_{ij} \partial_i \partial_j f_T(X).$$

In particular, for random vectors v with i.i.d. components the covariance matrix is constant D times the identity, and we obtain the usual heat equation

$$\partial_T f_T(X) = D\Delta_X f_T(X)$$

where D is called the *diffusion coefficient*.

2.2 Markov processes and their generators

Let X_t, $t \geq 0$, be a continuous-time stochastic process, i.e. X_t is a one-parameter family of random variables, the parameter is usually called time. The *state space* of the process is the space from where X_t takes its values, in our case $X_t \in \mathbb{R}^d$ or $\mathbb{Z}^d$. As usual, $\mathbf{E}$ will denote the expectation with respect to the distribution of the process.

Definition 2.4. *X_t is a Markov process if for any $t \geq \tau$*

$$Dist(X_t \mid \{X_s\}_{s\in[0,\tau]}) = Dist(X_t \mid X_\tau)$$

i.e. if the conditional distribution of X_t conditioned on the family of events X_s in times $s \in [0, \tau]$ is the same as conditioned only on the event at time τ.

In simple words it means that the state, X_t, of the system at time t depends on the past between times $[0, \tau]$ only via the state of the system at the last time of the past interval, τ. All necessary information about the past is condensed in the last moment.

The process depends on its initial value, X_0. Let $\mathbf{E}_x$ denote the expectation value assuming that the process started from the point $x \in \mathbb{R}^d$ or $\mathbb{Z}^d$, i.e.

$$\mathbf{E}_x \varphi(X_t) = \mathbf{E}\{\varphi(X_t) \mid X_0 = x\}.$$

Here and in the sequel φ will denote functions on the state space, these are also called *observables*. We will not specify exactly the space of observables.

Markov processes are best described by their generators:

Definition 2.5. *The generator of the Markov process X_t is an operator acting on the observables φ and it is given by*

$$(\mathcal{L}\varphi)(x) := \frac{d}{d\varepsilon}\Big|_{\varepsilon=0+0} \mathbf{E}_x\, \varphi(X_\varepsilon).$$

This definition is a bit lousy since the function space on which $\mathcal{L}$ act is not defined. Typically it is either the space of continuous functions or the space of L^2-functions on the state space, but for both concept an extra stucture – topology or measure – is needed on the state space. Moreover, the generator is typically an unbounded operator, not defined on the whole function space but only on a dense subset. For this lecture we will leave these technicalities aside, but when necessary, we will think of the state space $\mathbb{R}^d, \mathbb{Z}^d$ equipped with their natural measures and for the space of functions we consider will the L^2-space.

The key fact is that the generator tells us everything about the Markov process itself. Let us demonstrate it by answering two natural questions in terms of the generator.

Question 1: Starting from $X_0 = x$, what is the distribution of X_t?

Answer: Let φ be an observable (function on the state space) and define

$$f_t(x) := \mathbf{E}_x\, \varphi(X_t). \tag{2.6}$$

We wish to derive an evolution equation for f_t:

$$\partial_t f_t(x) = \lim_{\varepsilon\to 0} \frac{1}{\varepsilon} \mathbf{E}_x \Big[\mathbf{E}_{X_\varepsilon} \varphi(\widetilde{X}_t) - \mathbf{E}_x\, \varphi(X_t) \Big],$$

since by the Markov property

$$\mathbf{E}_x\, \varphi(X_{t+\varepsilon}) = \mathbf{E}_x \mathbf{E}_{X_\varepsilon}(\varphi(\widetilde{X}_t)),$$

where $\widetilde{X}_t$ is a new copy of the Markov process started at the (random) point X_ε. Thus

$$\partial_t f_t(x) = \lim_{\varepsilon\to 0} \frac{1}{\varepsilon} \mathbf{E}_x \Big[f_t(X_\varepsilon) - f_t(x) \Big] = \frac{\mathrm{d}}{\mathrm{d}\varepsilon}\Big|_{\varepsilon=0+0} \mathbf{E}_x\, f_t(X_\varepsilon) = (\mathcal{L} f_t)(x).$$

Therefore f_t, defined in (2.6), solves the initial value problem

$$\partial_t f_t = \mathcal{L} f_t \qquad \text{with} \quad f_0(x) = \varphi(x).$$

Formally one can write the solution as

$$f_t = e^{t\mathcal{L}}\varphi.$$

If the observable is a Dirac delta function at a fixed point y in the state space,

$$\varphi(x) = \delta_y(x),$$

then the solution to the above initial value problem is called the *transition kernel* of the Markov process and it is denoted by $p_t(x, y)$:

$$\partial_t p_t(\cdot, y) = \mathcal{L}p_t(\cdot, y) \qquad \text{with} \quad p_0(\cdot, y) = \delta_y.$$

The intuitive meaning of the transition kernel is

$$p_t(x, y)\mathrm{d}y := \text{Prob}\{\text{After time } t \text{ the process is at } y + \mathrm{d}y \text{ if it started at } x \text{ at time } 0\}.$$

Question 2: Suppose that the initial value of the process, $X_0 = x$, is distributed according to a density function $\psi(x)$ on the state space. What is the probability that after time t the process is at y?

Answer:

$$g_t(y) := \text{Prob}(X_t \text{ at } y \text{ after time } t) = \int \psi(x)p_t(x, y)\mathrm{d}x.$$

It is an exercise to check (at least formally), that g_t solves the following initial value problem

$$\partial_t g_t = \mathcal{L}^* g_t \qquad \text{with} \quad g_0 = \psi,$$

where $\mathcal{L}^*$ denotes the adjoint of $\mathcal{L}$ (with respect to the standard scalar product of the L^2 of the state space).

2.3 Wiener process and its generator

The Wiener process is the rigorous mathematical construction of the Brownian motion. There are various ways to introduce it, we choose the shortest (but not the most intuitive) definition.

First we need the concept of the *Gaussian process*. We recall that the centered Gaussian random variables have the remarkable property that all their moments are determined by the covariance matrix (a random variable is called *centered* if its expectation is zero, $\mathbf{E}\, X = 0$). If $(X_1, X_2, \ldots, X_{2k})$ is a centered Gaussian vector-valued random variable, then the higher moments can be computed by **Wick's formula** from the second moments (or *two-point correlations*)

$$\mathbf{E}\, X_1 X_2 \ldots X_{2k} = \sum_{\pi \text{ pairing}} \prod_{(i,j)\in\pi} \mathbf{E}\, X_i X_j, \tag{2.7}$$

where the summation is over all possible (unordered) pairings of the indices $\{1, 2, \ldots 2k\}$.

A stochastic process $X_t \in \mathbb{R}^d$ is called *Gaussian*, if any finite dimensional projection is a Gaussian vector-valued random variable, i.e. for any $t_1 < t_2 < \ldots < t_k$, the vector $(X_{t_1}, X_{t_2}, \ldots, X_{t_n}) \in \mathbb{R}^{dn}$ is a Gaussian random variable.

Definition 2.6. *A continuous Gaussian stochastic process* $W_t = (W_t^{(1)}, \dots, W_t^{(d)}) \in \mathbb{R}^d$, $t \geq 0$, *is called the d-dimensional (standard) Wiener process if it satisfies the following*

i) $W_0 = 0$

ii) $\mathbf{E}\, W_t = 0$

iii) $\mathbf{E}\, W_s^{(a)} W_t^{(b)} = \min\{s,t\}\delta_{ab}$.

From Wick's formula it is clear that all higher moments of W_t are determined. It is a fact that the above list of requirements determines the Wiener process uniquely. We can now extend the central limit theorem to processes.

Theorem 2.7. *Recall the conditions of Theorem 2.1. The stochastic process*

$$\widetilde{X}_T^\varepsilon := \varepsilon^{1/2} \sum_{i=1}^{[T\varepsilon^{-1}]} v_i$$

(defined already in (2.2)) converges in distribution (as a stochastic process) to the Wiener process

$$\widetilde{X}_T^\varepsilon \Longrightarrow W_T$$

as $\varepsilon \to 0$.

The proof of this theorem is not trivial. It is fairly easy to check that the moments converge, i.e. for any $k \in \mathbb{N}^d$ multiindex

$$\mathbf{E}\, [\widetilde{X}_T^\varepsilon]^k \to \mathbf{E}\, W_T^k$$

and the same holds even for different times:

$$\mathbf{E}\, \left[\widetilde{X}_{T_1}^\varepsilon\right]^{k_1} \left[\widetilde{X}_{T_2}^\varepsilon\right]^{k_2} \dots \left[\widetilde{X}_{T_m}^\varepsilon\right]^{k_m} \to \mathbf{E}\, W_{T_1}^{k_1} W_{T_2}^{k_2} \dots W_{T_m}^{k_m},$$

but this is not quite enough. The reason is that a continuous time process is a collection of uncountable many random variables and the joint probability measure on such a big space is not necessarily determined by finite dimensional projections. If the process has some additional regularity property, then yes. The correct notion is the *stochastic equicontinuity* (or Kolmogorov condition) which provides the necessary compactness (also called tightness in this context). We will not go into more details here, see any standard book on stochastic processes.

Theorem 2.8. *The Wiener process on* $\mathbb{R}^d$ *is Markovian with generator*

$$\mathcal{L} = \frac{1}{2}\Delta.$$

Idea of the proof. We work in $d = 1$ dimension for simplicity. The Markovity can be easily checked from the explicit formula for the correlation function. The most important ingredient is that the Wiener process has *independent increments:*

$$\mathbf{E}\, (W_t - W_s) W_u = 0 \tag{2.8}$$

if $u \le s \le t$; i.e. the increment in the time interval $[s,t]$ is independent of the past. The formula (2.8) is trivial to check from property (iii).

Now we compute the generator using Definition 2.5

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\Big|_{\varepsilon=0+0} \mathbf{E}_0\varphi(W_\varepsilon) = \frac{\mathrm{d}}{\mathrm{d}\varepsilon}\Big|_{\varepsilon=0+0} \mathbf{E}_0\left(\varphi(0) + \varphi'(0)W_\varepsilon + \frac{1}{2}\varphi''(0)W_\varepsilon^2 + \dots\right) = \frac{1}{2}\varphi''(0).$$

Here we used that $\mathbf{E}_0 W_\varepsilon = 0$, $\mathbf{E}_0 W_\varepsilon^2 = \varepsilon$ and the dots refer to irrelevant terms that are higher order in ε. $\square$

Finally we remark that one can easily define a Wiener process with any nontrivial (non-identity) diffusion coefficient matrix D as long as it is positive definite (just write $D = A^*A$ and apply the linear transformation $W \to AW$ to the standard Wiener process). The generator is then

$$\mathcal{L} = \frac{1}{2}\sum_{i,j=1}^{d} D_{ij}\partial_i\partial_j.$$

2.4 Random jump process on the sphere S^{d-1} and its generator

The state space of this process is the unit sphere $S = S^{d-1}$ in $\mathbb{R}^d$. We are given a function

$$\sigma(v,u) : S \times S \to \mathbb{R}_+$$

which is interpreted as jump rate from the point v to u.

The process will be parametrized by a continuous time t, but, unlike the Wiener process, its trajectories will not be continuous, rather piecewise constant with some jumps at a discrete set of times. A good intuitive picture is that the particle jumping on the unit sphere has a random clock (so-called *exponential clock*) in its pocket. The clock emits a signal with the standard exponential distribution, i.e.

$$\mathrm{Prob}\{\text{there is a signal at } t + \mathrm{d}t\} = e^{-t}\mathrm{d}t,$$

and when it ticks, the particle, being at location v, jumps to any other location $u \in S$ according to the distribution $u \to \sigma(v,u)$ (with the appropriate normalization):

$$\mathrm{Prob}\{\text{The particle from } v \text{ jumps to } u + \mathrm{d}u\} = \frac{\sigma(v,u)\mathrm{d}u}{\int \sigma(v,u)\mathrm{d}u}.$$

The transition of the process at an infinitesimal time increment is given by

$$v_{t+\varepsilon} = \begin{cases} u + \mathrm{d}u & \text{with probability} \quad \varepsilon\sigma(v_t,u)\mathrm{d}u \\ v_t & \text{with probability} \quad 1 - \varepsilon\int\sigma(v_t,u)\mathrm{d}u \end{cases} \tag{2.9}$$

up to terms of order $O(\varepsilon^2)$ as $\varepsilon \to 0$.

Thus the generator of the process can be computed from Definition 2.5. Let

$$f_t(v) := \mathbf{E}_v\varphi(v_t),$$

assuming that the process starts from some given point $v_0 = v$ and φ is an arbitrary observable. Then

$$\begin{aligned}\partial_t f_t(v) &= \lim_{\varepsilon\to 0+0} \frac{1}{\varepsilon}\mathbf{E}_v\Big(\mathbf{E}_{v_\varepsilon}\varphi(\widetilde{v}_t) - \mathbf{E}_v\varphi(\widetilde{v}_t)\Big)\\ &= \int \mathrm{d}u\ \sigma(v,u)\Big[\mathbf{E}_u\varphi(\widetilde{v}_t) - \mathbf{E}_v\varphi(\widetilde{v}_t)\Big]\\ &= \int \mathrm{d}u\ \sigma(v,u)\big[f_t(u) - f_t(v)\big],\end{aligned}$$

where we used the Markov property in the first line and the infinitesimal jump rate from (2.9) in the second line. Note that with probability $1-\varepsilon\sigma(v,u)$ we have $v_\varepsilon = v$. Thus the generator of the random jump process is

$$(\mathcal{L}f)(v) := \int \mathrm{d}u\ \sigma(v,u)\big[f(u) - f(v)\big] \tag{2.10}$$

Note that it has two terms; the first term is called the *gain term* the second one is the *loss term*. The corresponding evolution equation for the time dependent probability density of the jump process,

$$\partial_t f_t(v) = \int \mathrm{d}u\ \sigma(v,u)\big[f_t(u) - f_t(v)\big] \tag{2.11}$$

is called *linear Boltzmann equation in velocity space*.

The elements of the state space S will be interpreted as velocities of a moving particle undergoing fictitious random collisions. Only the velocities are recorded. A velocity distribution $f(v)$ is called *equilibrium distribution*, if $\mathcal{L}f \equiv 0$; it is fairly easy to see that, under some nondegeneracy condition on σ, the equilibrium exists uniquely.

3 Classical mechanics of a single particle

Classical mechanics of a single particle in d-dimensions is given by a Hamiltonian (energy function) defined on the phase space $\mathbb{R}^d \times \mathbb{R}^d$:

$$H(v,x) := \frac{1}{2}v^2 + U(x). \tag{3.1}$$

Here x is the position, v is the momentum coordinate. For most part of this notes we will not distinguish between momentum and velocity, since we almost always consider the standard kinetic energy $\frac{1}{2}v^2$.

The Hamiltonian equation of motions is the following set of $2d$ coupled first order differential equations:

$$\dot{x}(t) = \partial_v H = v \qquad \dot{v}(t) = -\partial_x H = -\nabla U(x). \tag{3.2}$$

For example, in case of the free evolution, when the potential is zero, $U \equiv 0$, we have

$$x(t) = x_0 + v_0 t \tag{3.3}$$

i.e. linear motion with a constant speed.

Instead of describing each point particle separately, we can think of a continuum of (non-interacting) single particles, described by a phase space density $f(x, v)$. This function is interpreted to give the number of particles with given velocity at a given position, more precisely

$$\int_\Delta f(x,v)\mathrm{d}x\mathrm{d}v = \text{Number of particles at } x \text{ with velocity } v \text{ such that } (x,v) \in \Delta.$$

Another interpretation of the phase space density picture is that a single particle with velocity v_0 at x_0 can be described by the measure

$$f(x, v) = \delta(x - x_0)\delta(v - v_0),$$

and it is then natural to consider more general measures as well.

The system evolves with time, so does its density, i.e. we consider the time dependent phase space densities, $f_t(x, v)$. For example in case of a single particle governed by (3.2) with initial condition $x(0) = x_0$ and $v(0) = v_0$, the evolution of the phase space density is given by

$$f_t(x, v) = \delta(x - x(t))\delta(v - v(t)),$$

where $(x(t), v(t))$ is the Hamiltonian trajectory computed from (3.2).

It is easy to check that if the point particle trajectories satisfy (3.2), then f_t satisfies the following evolution equation

$$(\partial_t + v \cdot \nabla_x) f_t(x, v) = \nabla U(x) \cdot \nabla_v f_t(x, v), \tag{3.4}$$

which is called the *Liouville equation*. The left hand side is called the *free streaming term*. The solution to

$$(\partial_t + v \cdot \nabla_x) f_t(x, v) = 0 \tag{3.5}$$

is given by the linear transport solution

$$f_t(x, v) = f_0(x - vt, v),$$

where f_0 is the phase space density at time zero. This corresponds to the free evolution (3.3) in the Hamiltonian particle representation.

3.1 The linear Boltzmann equation

The linear Boltzmann equation is a phenomenological combination of the free flight equation (3.5) and the jump process on the sphere of the velocity space (2.11)

$$(\partial_t + v \cdot \nabla_x) f_t(x, v) = \int \sigma(u, v)\big[f_t(x, u) - f_t(x, v)\big]\mathrm{d}u. \tag{3.6}$$

Note that we actually have the adjoint of the jump process (observe that u and v are interchanged compared with (2.11)), so the solution determines:

$$f_t(x, v) = \text{Prob}\{\text{the process is at } (x, v) \text{ at time } t\},$$

given the initial probability density $f_0(x, v)$ (Question 2 in Section 2.2).

The requirement that the velocities are constrained to the sphere corresponds to energy conservation. Of course there is no real Hamiltonian dynamics behind this equation: the jumps are stochastic.

Notice that the free flight plus collision process standing behind the Boltzmann equation is actually a random walk in continuous time. In the standard random walk the particle "jumps" in a new random direction after every unit time. In the Boltzmann process the jumps occur at random times ("exponential clock"), but the effect on a long time scale is the same. In particular, the long time evolution of the linear Boltzmann equation is diffusion (Wiener process) in position space.

The following theorem formulates this fact more precisely. We recall that to translate the velocity process into position space, one has to integrate the velocity, i.e. will consider

$$x_t = \int_0^t v_s \mathrm{d}s.$$

Theorem 3.1. *Let v_t be a random velocity jump process given by the generator* (2.11)*. Then*

$$X_\varepsilon(T) =: \varepsilon^{1/2} \int_0^{T/\varepsilon} v_t \, \mathrm{d}t \to W_T \quad \textit{(in distribution,)}$$

where W_T is a Wiener process with diffusion coeffient being the velocity autocorrelation

$$D = \int_0^\infty R(t)\mathrm{d}t, \qquad R(t) := \mathbf{E}\, v_0 v_t.$$

Here $\mathbf{E}$ *is with respect to the equilibrium measure of the jump process on S^{d-1}.*

4 Quantum mechanics of a single particle

4.1 Wavefunction, Wigner transform

The state space of a quantum particle in d-dimensions is $L^2(\mathbb{R}^d)$ or $L^2(\mathbb{Z}^d)$. Its elements are called L^2-wavefunctions and they are usually denoted by $\psi = \psi(x) \in L^2(\mathbb{R}^d)$ or $L^2(\mathbb{Z}^d)$. We assume the normalization, $\|\psi\|_2 = 1$. We recall the interpretation of the wave function: the position space density $|\psi(x)|^2 \mathrm{d}x$ gives the probability of finding an electron at a spatial domain:

$$\int_\Delta |\psi(x)|^2 \mathrm{d}x = \mathrm{Prob}\{\text{the particle's position is in } \Delta\}.$$

Similarly, the Fourier transform of ψ,

$$\widehat{\psi}(v) := \int_{\mathbb{R}^d} e^{-iv\cdot x} \psi(x) \mathrm{d}x,$$

determines the momentum space density:

$$\int_\Delta |\widehat{\psi}(v)|^2 \mathrm{d}v = \mathrm{Prob}\{\text{the particle's momentum is in } \Delta\}.$$

In the lattice case, i.e. when $\mathbb{Z}^d$ replaces $\mathbb{R}^d$ as the physical space, the Fourier transform is replaced with Fourier series. In these notes we neglect all 2π's that otherwise enter the definition of the Fourier transform.

By the Heisenberg uncertainty principle, one cannot simultaneously determine the position and the momentum of a quantum particle, thus the concept of classical phase space density does not generalize directly to quantum mechanics. Nevertheless one can define a substitute for it, namely the *Wigner transform*. For any L^2-wavefunction ψ we define the Wigner transform of ψ as

$$W_\psi(x,v) := \int \overline{\psi}\Big(x+\frac{z}{2}\Big)\psi\Big(x-\frac{z}{2}\Big)e^{ivz}\mathrm{d}z,$$

and we still interpret it as "quantum phase space density".

It is easy to check that W_ψ is always real but in general is not positive (thus it cannot be the density of a positive measure – in coincidence with the Heisenberg principle). However, its marginals reconstruct the position and momentum space densities, as the following formulas can be easily checked:

$$\int W_\psi(x,v)\mathrm{d}v = |\psi(x)|^2, \qquad \int W_\psi(x,v)\mathrm{d}x = |\widehat{\psi}(v)|^2.$$

In particular, for normalized wave functions $\|\psi\|_2 = 1$ we have

$$\iint W_\psi(x,v)\mathrm{d}v\mathrm{d}x = 1. \tag{4.1}$$

We remark, that for the lattice case some care is needed for the proper definition of the Wigner transform, since $x \pm \frac{z}{2}$ may not be a lattice site. The precise definition in this case is

$$W_\psi(x,v) := \sum_{\substack{y,z\in\mathbb{Z}^d \\ y+z=2x}} e^{iv(y-z)}\bar{\psi}(y)\psi(z), \tag{4.2}$$

where $\psi \in \ell^2(\mathbb{Z}^d)$, i.e. $y,z \in \mathbb{Z}^d$, but $x \in \big(\mathbb{Z}/2\big)^d$. The formulas for marginal of the Wigner transform modify as follows:

$$\int W_\psi(x,v)\mathrm{d}v = 2^d|\psi(x)|^d$$

if $x \in \mathbb{Z}^d$ and it is zero if $x \in \big(\mathbb{Z}/2\big)^d \setminus \mathbb{Z}^d$. We still have

$$\int_{\left(\mathbb{Z}/2\right)^d} W_\psi(x,v)\mathrm{d}x = 2^{-d}\sum_{x\in\left(\mathbb{Z}/2\right)^d} W_\psi(x,v) = |\widehat{\psi}(v)|^2$$

and

$$\int_{\left(\mathbb{Z}/2\right)^d} \mathrm{d}x \int \mathrm{d}v W_\psi(x,v) = \|\psi\|^2.$$

Often we will use the Wigner transform in the Fourier representation, by which we will always mean Fourier transform in the first (x) variable only, i.e. with the convention

$$\widehat{f}(\xi) = \int e^{-ix\xi}f(x)\mathrm{d}x,$$

we define

$$\widehat{W}_\psi(\xi, v) := \int e^{-ix\xi} W_\psi(x, v)\mathrm{d}x.$$

After a simple calculation, we have

$$\widehat{W}_\psi(\xi, v) = \overline{\widehat{\psi}\Big(v - \frac{\xi}{2}\Big)}\widehat{\psi}\Big(v + \frac{\xi}{2}\Big). \tag{4.3}$$

In the discrete case we have $\xi \in R^d/(2 \cdot 2\pi\mathbb{Z}^d)$ and

$$\widehat{W}(\xi, v) = \overline{\widehat{\psi}\Big(v - \frac{\xi}{2}\Big)}\widehat{\psi}\Big(v + \frac{\xi}{2}\Big).$$

More generally, if $J(x, v)$ is a classical phase space observable, the scalar product

$$\langle J, W_\psi\rangle = \int J(x, v)W_\psi(x, v)\mathrm{d}x\mathrm{d}v$$

can be interpreted as the expected value of J in state ψ. Recall that "honest" quantum mechanical observables are self-adjoint operators $\mathcal{O}$ on $L^2(\mathbb{R}^d)$ and their expected value is given by

$$\langle\psi, \mathcal{O}\psi\rangle = \int \overline{\psi}(x)\ (\mathcal{O}\psi)(x)\mathrm{d}x.$$

For a large class of observables there is a natural relation between observables $\mathcal{O}$ and their phase space representations (called *symbols*) that are functions on the phase space like $J(x, v)$. For example, if J depends only on x or only on v, then the corresponding operator is just the standard quantization of J, i.e.

$$\int J(x)W_\psi(x, v)\mathrm{d}x\mathrm{d}v = \langle\psi, J\psi\rangle$$

where J is a multiplication operator on the right hand side, or

$$\int \overline{J}(v)W_\psi(x, v)\mathrm{d}x\mathrm{d}v = \langle\psi, J(-i\nabla)\psi\rangle,$$

and similar relations hold for the Weyl quantization of any symbol $J(x, v)$.

We also remark that the map $\psi \to W_\psi$ is invertible, i.e. one can fully reconstruct the wave function from its Wigner transform. On the other hand, not every real function of two variables (x, v) is the Wigner transform of some wavefunction.

4.2 Hamilton operator and the Schrödinger equation

The quantum dynamics is generated by the *Hamilton operator*

$$H = -\frac{1}{2}\Delta_x + U(x)$$

acting on $\psi \in L^2(\mathbb{R}^d)$. The first term is interpreted as the kinetic energy and it is the quantization of the classical kinetic energy $\frac{1}{2}v^2$ (compare with (3.1)). The momentum operator is

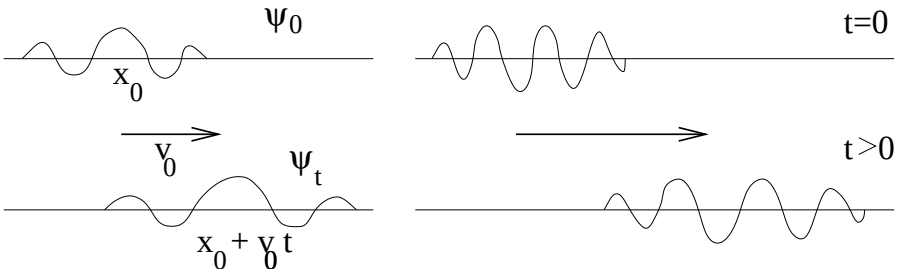

Figure 4: Evolution of slower (left) and a faster (right) wave packet

$p = -i\nabla$ and we set the mass equal one, so momentum and velocity coincide. [To be pedantic, we should use the word momentum everywhere instead of velocity; since the momentum is the canonical quantity, then the dispersion relation, $e(p)$, defines the kinetic energy as a function of the momentum operator, and its gradient, $v = \nabla e(p)$ is the physical velocity. Since we will mostly use the dispersion relation $e(p) = \frac{1}{2}p^2$, the momentum and velocity coincide. However, for the lattice model the velocity and the momentum will differ.]

The evolution of a state is given by the Schrödinger equation

$$i\partial_t \psi_t = H\psi_t = \Big(-\frac{1}{2}\Delta + U \Big)\psi_t$$

with a given initial data $\psi_{t=0} = \psi_0$. Formally, the solution is

$$\psi_t = e^{-itH}\psi_0.$$

If H is self-adjoint in L^2, then the unitary group e^{-itH} can be defined by spectral theorem. Almost all Hamilton operators in mathematical physics are self-adjoint, however the self-adjointness usually requires some proof. We will neglect this issue here, but we only mention that self-adjointness is more than the symmetry of the operator, because H is typically unbounded when issues about the proper domain of the operator become relevant.

Note the complex i in the Schrödinger equation, it plays an absolutely crucial role. It is responsible for the wave-like character of quantum mechanics. The quantum evolution are governed by *phase* and *dispersion*. Instead of giving precise explanations, look at Fig. 4: the faster the wave oscillates, the faster it moves.

We can also justify this picture by some sketchy calculation (all these can be made rigorous in the so-called semiclassical regime). We demonstrate that the free ($U \equiv 0$) evolution of

$$\psi_0(x) := e^{iv_0 x} A(x - x_0)$$

(with some fixed x_0, v_0) after time t is supported around $x_0 + v_0 t$. Here we tacitly assume that A is not an oscillatory function (e.g. positive) and the only oscillation is given explicitly in the phase $e^{iv_0 x}$, mimicking a plane wave.

We compute the evolution of ψ_0:

$$\psi_t(x) = \int e^{iv(x-y)} e^{-itv^2/2} \underbrace{e^{iv_0 y} A(y - x_0)}_{=:\psi_0(y)} \, \mathrm{d}y\mathrm{d}v$$

$$= e^{iv_0 x_0} \int e^{iv(x-x_0)} e^{-itv^2/2} \widehat{A}(v - v_0) \, \mathrm{d}v \sim \int e^{i\Phi(v)} \widehat{A}(v - v_0) \, \mathrm{d}v$$

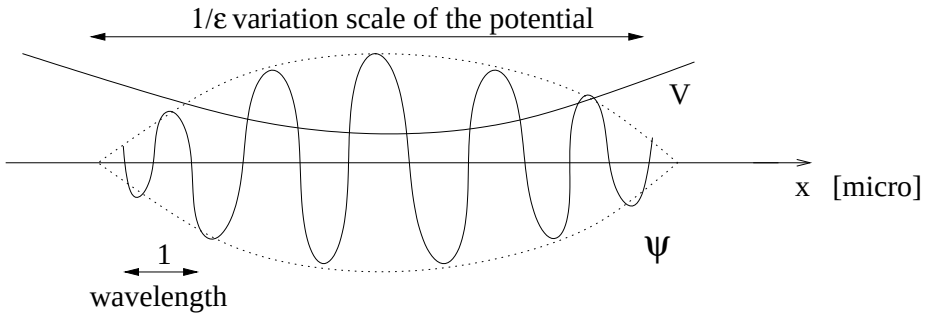

Figure 5: Semiclassical setup: short wavelength, large scale envelope and potential

with a phase factor

$$\Phi(v) := v(x - x_0) - \frac{1}{2}tv^2.$$

We apply the stationary phase argument: the integral is concentrated around the stationary points of the phase. Thus

$$0 = \nabla_v \Phi = x - x_0 - tv$$

gives $x = x_0 + vt \approx x_0 + v_0 t$ if $\widehat{A}$ is sufficiently localized.

This argument can be made rigorous if a scale separation ansatz is used. We introduce a small parameter ε and assume that

$$\psi_0(x) = \varepsilon^{d/2} e^{iv_0 x} A(\ \varepsilon(x - x_0)\ ),$$

i.e. the wave has a slowly varying envelope (amplitude) and fast phase. The prefactor is chosen to keep the normalization $\|\psi_0\| = 1$ independently of ε. Such states and their generalizations of the form

$$\varepsilon^{d/2} e^{iS(\varepsilon x)/\varepsilon} A(\ \varepsilon(x - x_0)\ ) \tag{4.4}$$

are called *WKB-states*. These states are characterized by a slowly varying amplitude and wavelength profile and by a fast oscillation. Similar picture holds if the potential is nonzero, but slowly varying, i.e. $U(x)$ is replaced with $U(\varepsilon x)$.

4.3 Semiclassics

The WKB states and the rescaled potential belong to the *semiclassical theory* of quantum mechanics. A quantum mechanical system is in the *semiclassical regime* whenever all the data (potential, magnetic field, initial profile etc.) live on a much bigger scale than the quantum wavelength.

The scale separation parameter is usually explicitly included in the Schrödinger equation, for example:

$$i\partial_t \psi_t(x) = \Big[-\frac{1}{2}\Delta_x + U(\varepsilon x) \Big] \psi_t(x). \tag{4.5}$$

This equation is written in microscopic coordinates. It becomes more familiar if we rewrite it in macroscopic coordinates

$$(X, T) = (x\varepsilon, t\varepsilon),$$

where (4.5) takes the form

$$i\varepsilon\partial_T\Psi_T(X) = \Big[-\frac{\varepsilon^2}{2}\Delta_X + U(X)\Big]\Psi_T(X).$$

Traditionally ε is denoted h in this equation and is referred to the (effective) Planck constant, whose smallness in standard units justifies taking the limit $h \to 0$ (the so-called *semiclassical limit*). No matter in what coordinates and units, the essential signature of the semiclassical regime is the scale separation between the wavelength and any other lengthscales.

Wigner transform $W_\psi(x, v)$, written originally in microscopic coordinates, can also be rescaled:

$$W^\varepsilon_\psi(X, V) := \varepsilon^{-d} W_\psi\Big(\frac{X}{\varepsilon}, V\Big).$$

Note that apart from the rescaling of the X variable (V variable is unscaled!), a normalizing prefactor ε^{-d} is added to keep the integral of W^ε normalized (see (4.1))

$$\iint W^\varepsilon_\psi(X, V)\mathrm{d}X\mathrm{d}V = 1.$$

The following theorem is a simple prototype of the theorems we wish to derive in more complicated situations. It shows how the quantum mechanical Schrödinger evolution can be approximated by a classical equation in a certain limiting regime:

Theorem 4.1. *Consider a sequence of initial states ψ^ε_0 whose rescaled Wigner transform has a weak limit:*

$$\lim_{\varepsilon\to 0} W^\varepsilon_{\psi^\varepsilon_0}(X, V) \rightharpoonup W_0(X, V).$$

For example, the WKB states given in (4.4) have a weak limit

$$W_0(X, V) = |A(X)|^2\delta(V - \nabla S(X)).$$

Let ψ^ε_t denote the solution to the semiclassical Schrödinger equation (4.5) with initial state ψ^ε_0. Then the weak limit of the rescaled Wigner transform of the solution at time $t = T/\varepsilon$,

$$W_T(X, V) := \lim_{\varepsilon\to 0} W^\varepsilon_{\psi^\varepsilon_{T/\varepsilon}}(X, V)$$

satisfies the Liouville equation:

$$(\partial_T + V\cdot\nabla_X)W_T(X, V) = \nabla U(X)\cdot\nabla_V W_T(X, V). \tag{4.6}$$

Recall that the Liouville equation is equivalent to the classical Hamilton dynamics (see the derivation of (3.4)).

In this way, the Liouville equation (4.6) mimics the semiclassical quantum evolution on large scales (for a proof, see e.g. [40]). Notice that weak limit is taken; this means that the small scale structure of the Wigner transform $W_{\psi^\varepsilon_t}$ is lost. Before the weak limit, the Wigner transform of ψ^ε_t carries all information about ψ^ε_t, but the weak limit means testing it against functions that live on the macroscopic scale. This phenomenon will later be essential to understand why irreversibility of the Boltzmann equation does not contradict the reversibility of Schrödinger dynamics.

5 Random Schrödinger operators

5.1 Quantum Lorentz gas or the Anderson model

Now we introduce our basic model, the quantum Lorentz gas. Physically, we wish to model electron transport in a disordered medium. We assume that electrons are non-interacting, thus we can focus on the evolution of a single electron. The disordered medium is modelled by a potential describing random impurities (other models are also possible, e.g. one could include random perturbations in the kinetic energy term). This can be viewed as the quantum analogue of the classical Lorentz gas, where a classical particle is moving among random obstacles.

Definition 5.1. *The model of a single particle described by the Schrödinger equation on* $\mathbb{R}^d$ *or on* $\mathbb{Z}^d$*,*

$$i\partial_t \psi_t(x) = H\psi_t(x), \qquad H = -\Delta_x + \lambda V_\omega(x),$$

is called the **quantum Lorentz gas**, *if* $V_\omega(x)$ *is a random potential at location* x*.*

Notation: The subscript ω indicates that $V(x)$ is a random variable that also depends on an element ω in the probability space.

More generally, we can consider

$$H = H_0 + \lambda V, \tag{5.1}$$

where H_0 is a deterministic Hamiltonian, typically $H_0 = -\Delta$. We remark that a periodic background potential can be added, i.e.

$$H = -\Delta_x + U_{per}(x) + \lambda V_\omega(x)$$

can be investigated with similar methods. We recall that the operator with a periodic potential and no randomness,

$$H_0 = -\Delta_x + U_{per}(x),$$

can be more or less explicitly diagonalized by using the theory of Bloch waves. The transport properties are similar to those of the free Laplacian. This is called the *periodic quantum Lorentz gas* and it is the quantum analogue of the Sinai billiard. However, while the dynamics of the Sinai billiard has a very complicated structure, its quantum version is quite straightforward (the theory of Bloch waves is fairly simple).

We also remark that for the physical content of the model, it is practically unimportant whether we work on $\mathbb{R}^d$ or on its lattice approximation $\mathbb{Z}^d$. The reason is that we are investigating long time, large distance phenomenon; the short scale structure of the space does not matter. However, technically $\mathbb{Z}^d$ is much harder (a bit unusual, since typically $\mathbb{R}^d$ is harder as one has to deal with the ultraviolet regime). If one works on $\mathbb{Z}^d$, then the Laplace operator is interpreted as the discrete Laplace operator on $\mathbb{Z}^d$, i.e.

$$(\Delta f)(x) := 2d\, f(x) - \sum_{|e|=1} f(x+e). \tag{5.2}$$

In Fourier space this corresponds to the dispersion relation

$$e(p) = \sum_{j=1}^{d} (1 - \cos p^{(j)}) \tag{5.3}$$

on the torus $p \in [-\pi, \pi]^d$.

The random potential can be fairly arbitrary, the only important requirement is that it cannot have a long-distance correlation in space. For convenience, we will assume i.i.d. random potential for the lattice case with a standard normalization:

$$\{V(x) \ : \ x \in \mathbb{Z}^d\} \quad \text{i.i.d} \qquad \mathbf{E}\, V(x) = 0, \ \ \mathbf{E}\, V^2(x) = 1.$$

The coupling constant λ can then be used to adjust the strength of the randomness. We can also write this random potential as

$$V(x) = \sum_{\alpha \in \mathbb{Z}^d} v_\alpha \delta(x - \alpha), \tag{5.4}$$

where $\{v_\alpha \ : \ \alpha \in \mathbb{Z}^d\}$ is a collection of i.i.d. random variables and δ is the usual lattice delta function.

For continuous models, the random potential can, for example, be given as follows;

$$V(x) = \sum_{\alpha \in \mathbb{Z}^d} v_\alpha B(x - \alpha),$$

where B is a nice (smooth, compactly supported) single site potential profile and $\{v_\alpha \ : \ \alpha \in \mathbb{Z}^d\}$ is a collection of i.i.d. random variables. It is also possible to let the randomness perturb the location of the obstacles instead their strength, i.e.

$$V(x) = \sum_{\alpha \in \mathbb{Z}^d} B(x - y_\alpha(\omega)),$$

where, for example, $y_\alpha(\omega)$ is a random point in the unit cell around $\alpha \in \mathbb{Z}^d$, or even more canonically, the collection $\{y_\alpha(\omega)\}_\alpha$ is just a Poisson point process in $\mathbb{R}^d$. The combination of these two models is also possible and meaningful, actually in our work [24] we consider

$$V(x) = \int B(x - y) \mathrm{d}\mu_\omega(y) \tag{5.5}$$

where μ_ω is a Poisson point process with homogeneous unit density and with i.i.d. coupling constants (weights), i.e.

$$\mu_\omega = \sum_\alpha v_\alpha(\omega) \delta_{y_\alpha(\omega)}, \tag{5.6}$$

where $\{y_\alpha(\omega) \ : \ \alpha = 1, 2, \ldots\}$ is a realization of the Poission point process and $\{v_\alpha\}$ are independent (from each other and from y_α as well) real valued random variables.

The lattice model $-\Delta + \lambda V$ with i.i.d. random potentials (5.4) is called *Anderson model*. It was invented by Anderson [4] who was the first to realize that electrons move quite differently

in disordered media than in free space or in a periodic background. The main phenomenon Anderson discovered was the *Anderson localization*, asserting that at sufficiently strong disorder (or in $d = 1, 2$ at any nonzero disorder) the electron transport stops. We will explain this in more details later, here we just remark that Anderson was awarded the Nobel Prize in 1977 for this work.

However, in $d \geq 3$ dimensions electron transport is expected despite the disorder if the disorder is weak. This is actually what we experience in real life; the eletric wire is conducting although it does not have a perfect periodic lattice structure. However, the nature of the electric transport changes: we expect that on large scales it can be described by a diffusive equation (like the heat equation) instead of a wave-type equation (like the Schrödinger equation).

Note that the free Schrödinger equation is ballistic (due to wave coherence). This means that if we define the *mean square displacement* by the expectation value of the observable x^2 at state ψ_t,

$$\langle\, x^2 \,\rangle_t := \int \mathrm{d}x |\psi_t(x)|^2 x^2 \qquad \Big(= \int \mathrm{d}x \big| e^{-itH} \psi_0(x) \big|^2 x^2 \Big), \tag{5.7}$$

then it behaves as

$$\langle\, x^2 \,\rangle_t \sim t^2$$

for large $t \gg 1$ as it can be easily computed. In contrast, the mean square displacement for the heat equation scales as t and not as t^2 (see (2.4)).

Thus the long time transport of the free Schrödinger equation is very different from the heat equation. We nevertheless claim, that in a weakly disordered medium, the long time Schrödinger evolution can be described by a diffusion (heat) equation. Our results answer to the intriguing question how the diffusive character emerges from a wave-type equation.

5.2 Known results about the Anderson model

We will not give a full account of all known mathematical results on the Anderson model, we just mention the main points. The main issue is to study the dichotomic nature of the Anderson model, namely that at low dimension ($d = 1, 2$) or at high disorder ($\lambda \geq \lambda_0(d)$) or near the spectral edges of the unperturbed operator H_0 the time evolved state remains localized, while at high dimension ($d \geq 3$), at low disorder and away from the spectral edges it is delocalized.

There are several signatures of (de)localization and each of them can be used for rigorous definition. We list three approaches:

i) *Spectral approach.* If H has pure point (PP) spectrum then the system is in the localized regime, if H has absolutely continuous (AC) spectrum then it is in the delocalized regime. (The singular continuous spectrum, if appears at all, gives rise to anomalous diffusions). It is to be noted that even in the pure point regime the spectrum is dense.

ii) *Dynamical approach.* One considers the mean square displacement (5.7) and the system is in the localized regime if

$$\sup_{t \geq 0} \langle x^2 \rangle_t < \infty$$

(other observables can also be considered).

iii) *Conductor or Insulator?* This is the approach closest to physics and mathematically it has not been sufficiently elaborated (for a physical review, see [37]). In this approach one imposes an external voltage to the system and computes the currect and the ohmic resistance.

These criteria to distinguish between localized and delocalized regimes are not fully equivalent (especially in $d = 1$ dimensional systems there are many anomalous phenomena due to the singular continuous spectrum), but we do not go into more details.

The mathematically most studied approach is the spectral method. The main results are the following:

i) In $d = 1$ dimensions all eigenfunctions are localized for all $\lambda \neq 0$. (Goldsheid, Molchanov and Pastur [31])

ii) In $d \geq 1$ localization occurs for large λ or for energies near the edges of the spectrum of H_0 (see (5.1)). This was first proven by the groundbreaking work of Fröhlich and Spencer [28] on the exponential decay of the resolvent via the *multiscale method*. (the absence of AC spectrum was then proved by Fröhlich, Martinelli, Spencer and Scoppola [27] and the exponential localization by Simon and Wolff [43]). Later a different method was found by Aizenman and Molchanov [2] (*fractional power method*). The spectrum is not only pure point, but the eigenfunctions decay exponentially (*strong localization*).

iii) Many more localization results were obtained by variants of these methods for various models, including magnetic field, acoustic waves, localized states along sample edges etc. We will not attempt to give a list of references here.

Common in all localization results is that the random potential dominates, i.e. in one way or another the free evolution H_0 is treated as a perturbation.

It is important to understand that the transport in the free Schrödinger equation is due to the coherence of the travelling wave. Random potential destroys this coherence; the stronger the randomness is, the surer is the destruction. This picture is essential to understand why random potentials may lead to localization even at energies that belong to the regime of classical transport. For example in $d = 1$ any small randomness stops transport, although the classical dynamics is still ballistic at energies that are higher than the maximal peak of the potential. In other words, Anderson localization is a truly quantum phenomenon, it cannot be explained by a heuristics based upon classically trapped particles.

5.3 Major open conjectures about the Anderson model

All the previous results were for the localization regime, where one of the two main methods (multiscale analysis or fractional power method) is applicable. The complementary regimes remain unproven. Most notably is the following list of open questions and conjectures:

i) [**Extended states conjecture**] For small $\lambda \leq \lambda_0(d)$ and in dimension $d \geq 3$ the spectrum is absolutely continuous away from the edges of H_0. In particular, there exists

a threshold value (called *mobility edge*) near the edges of the unperturbed spectrum that separates the two spectral types.

This conjecture has been proven only in the following three cases:

a) Bethe lattice (infinite binary tree) that roughly corresponds to $d = \infty$ (Klein [34], recently different proofs were obtained in [3] and [26]).

b) Sufficiently decaying random potential, $\mathbf{E}\, V_x^2 = |x|^{-\alpha}$, for $\alpha > 1$ (Bourgain [7]). The decay is sufficiently strong such that the typical number of collisions with obstacles is finite. Note that the randomness is not stationary.

c) In $d = 2$ with a constant magnetic field the spectrum cannot be everywhere pure point. (Germinet, Klein and Schenker [29])

ii) [**Two dimensional localization**] In $d = 2$ dimensions all eigenfunctions are localized for all λ. (i.e. the model in $d = 2$ dimensions behaves as in $d = 1$).

iii) [**Quantum Brownian motion conjecture**] For small λ and $d \geq 3$, the location of the electron is governed by a heat equation in a vague sense:

$$\partial_t |\psi_t(x)|^2 \sim \Delta_x |\psi_t(x)|^2 \qquad \Longrightarrow \langle\, x^2 \,\rangle_t \sim t \qquad t \gg 1. \tag{5.8}$$

The precise formulation of the first statement requires a scaling limit. The second statement about the diffusive mean square displacement is mathematically precise, but what really stands behind it is a diffusive equation that on large scales mimics the Schrödinger evolution. Moreover, the dynamics of the quantum particle converges to the Brownian motion as a process as well; this means that the joint distribution of the quantum densities $|\psi_t(x)|^2$ at different times $t_1 < t_2 < \ldots < t_n$ converges to the corresponding finite dimensional marginals of the Wiener process.

Note that the "Quantum Brownian motion conjecture" is much stronger than the "Extended states conjecture", since the former more precisely describes how the states extend. All these three open conjectures have been outstanding for many years and we seem to be far from their complete solutions.

Fig. 6 depicts the expected phase diagram of the Anderson model in dimensions $d \geq 3$. The picture shows the different spectral types (on the horizontal energy axis) at various disorder. The grey regimes indicate what has actually been proven.

Our main result, explained in the next sections, is that the "Quantum Brownian motion conjecture" in the sense of (5.8) holds *in the scaling limit* up to times $t \sim \lambda^{-2-\kappa}$. More precisely, we will fix a positive number κ and we will consider the family of random Hamiltonians

$$H = H_\lambda = -\frac{1}{2}\Delta + \lambda V$$

parametrized by λ. We consider their long time evolution up to times $t \sim \lambda^{-2-\kappa}$ and we take $\lambda \to 0$ limit. We show that, after appropriately rescling the space and time, the heat equation emerges. Note that the time scale depends on the coupling parameter. This result is of course

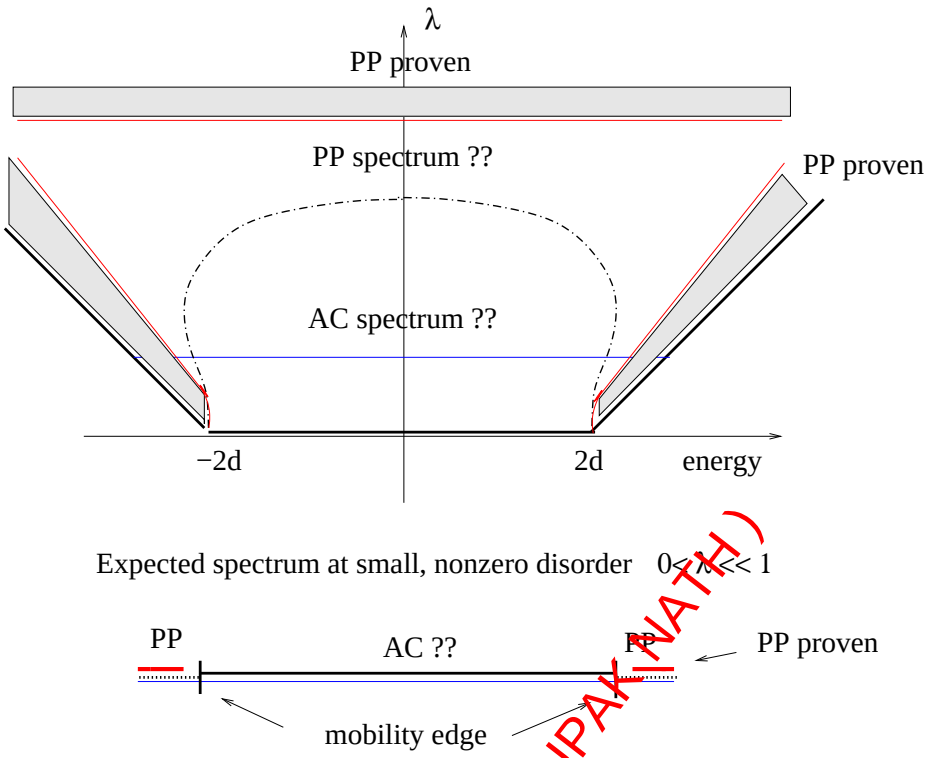

Figure 6: Phase diagram of the Anderson model in $d = 3$

far from either of the conjectures i) and ii) above, since those conjectures require fixing λ (maybe small but fixed) and letting $t \to \infty$.

We emphasize that the typical number of collisions is $\lambda^2 t$. The reason is that the quantum rate of collision with a potential that scales as $O(\lambda)$ is of order λ^2. This follows from simple scattering theory: if

$$\widetilde{H} = -\Delta + \lambda V_0$$

is a Hamiltonian with a single bump potential (i.e. V_0 is smooth, compactly supported) and ψ_{in} denotes the incoming wave, then after scattering the wave splits into two parts;

$$e^{-it\widetilde{H}}\psi_{in} = \beta e^{it\Delta}\psi_{in} + \psi_{sc}(t), \qquad t \gg 1. \tag{5.9}$$

Here $\psi_{sc}(t)$ is the scattered wave (typically spherical if V_0 is spherical) while the first term describes the transmitted wave that did not notice the obstacle (apart from its amplitude is scaled down by a factor β). Elementary calculation then shows that the scattered and transmitted waves are (almost) orthogonal and their amplitudes satisfy

$$\|\psi_{sc}(t)\|^2 = O(\lambda^2), \qquad \beta^2 = 1 - O(\lambda^2). \tag{5.10}$$

Therefore the incoming wave scatters with a probability $O(\lambda^2)$.

Thus up to time t, the particle encounters $\lambda^2 t$ collisions. In our scaling

$$n := \text{Number of collisions} \sim \lambda^2 t \sim \lambda^{-\kappa} \to \infty.$$

The asymptotically infinite number of collisions is necessary to detect the diffusion (Brownian motion, heat equation), similarly as the Brownian motion arises from an increasing number of independent "kicks" (steps in random walk, see Theorem 2.7).

6 Main result

In this section we formulate our main result precisely, i.e. what we mean by that "Quantum Brownian motion conjecture" holds in the scaling limit up to times $t \sim \lambda^{-2-\kappa}$. Before going into the precise formulation or the sketch of any proof, we will explain why it is a difficult problem.

6.1 Why is this problem difficult?

The dynamics of a single quantum particle among random scatterers (obstacles) is a multiple scattering process. The quantum wave function bumps into an obstacle along its evolution, and according to standard scattering theory, it decomposes into two parts: a wave that goes through the obstacle unnoticing it and a scattering wave (5.9). The scattered wave later bumps into another obstacle etc, giving rise to a complicated multiple scattering picture, similar to the one on Fig. 7. Obviously, the picture becomes more and more complicated as the number of collisions, n, increases. Finally, when we perform a measurement, we select a domain in space and we compute the wave function at the point. By the superposition principle of quantum mechanics, at that point *all* elementary scattering waves have to be added up; and eventually there exponentially many of them (in the parameter λ^{-1}). They are quantum waves, i.e. complex valued functions, so they must be added together with their phases. The cancellations in this sum due to the phases are essential even to get the right order of magnitude. Thus one way or another one must trace all elementary scattering waves with an enormous precision (the wavelength is order one in microscopic units but the waves travel at distance λ^{-2} between two collisions!)

It is important to emphasize that this system is **not semiclassical**. Despite the appearance of a small parameter, what matters is that the typical wavelength remains comparable with the spatial variation lengthscale of the potential. Thus the process remains quantum mechanical and cannot be described by any semiclassical method.

6.2 Scales

Since the collision rate is $O(\lambda^2)$, and the typical velocity of the particle is unscaled (order 1), the typical distance between two collisions (the so-called *mean free path*) is $L = O(\lambda^{-2})$ and the time elapsed between consecutive collisions is also $O(\lambda^{-2})$.

6.2.1 Kinetic scale

In order to see some nontrivial effects of the collisions, we have to run the dynamics at least up to times of order λ^{-2}. Within that time, the typical distance covered is also of order

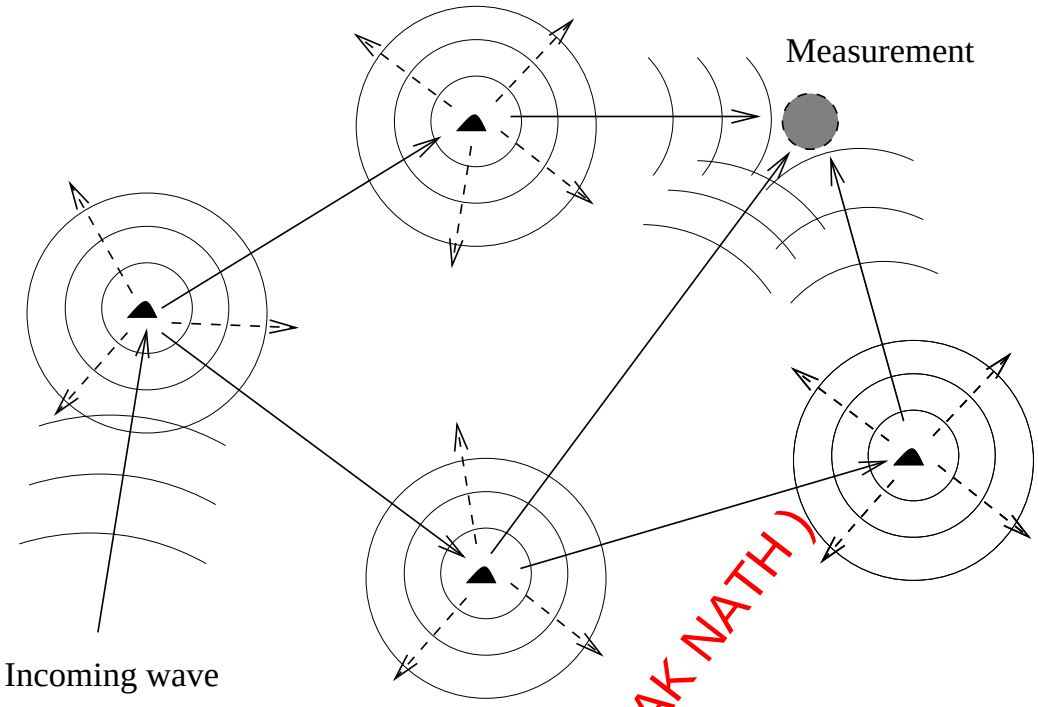

Figure 7: Schematic picture of multiple scattering

λ^{-2} (always understood in microscopic coordinates). Thus the simplest scale that yields a nontrivial result is the so-called **kinetic scaling**,

$$t = \lambda^{-2}T, \qquad x = \lambda^{-2}X, \qquad n = \lambda^2 t = O(1),$$

i.e. the space and time are scaled by λ^{-2} and the typical number of collisions $\lambda^2 t$ is of order 1.

One convenient way to quantify Fig. 7 is that one works in a large quadratic box whose size is at least comparable with the mean free path (otherwise the boundary would have a too big effect). Thus one can consider the box

$$\Lambda = [0, L]^d$$

and put $|\Lambda| = L^d$ obstacles in it, e.g. one at every lattice site. For definiteness, we work with the Anderson model, i.e.

$$H = -\Delta_x + \lambda \sum_{\alpha \in \mathbb{Z}^d} v_\alpha \delta(x - \alpha).$$

Each elementary scattering wave corresponds to a particular (ordered) sequence of random obstacles. Since the total number of obstacles is L^d, after n collisions we have $\sim |\Lambda|^n$ elementary waves to sum up in Fig. 7, thus

$$\psi_t = \sum_A \psi_A, \tag{6.1}$$

where $A = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ is a sequence of obstacle locations ($\alpha_i \in \mathbb{Z}^d$) and ψ_A describes the elementary scattering wave that has undergone the collisions with obstacles at $\alpha_1, \alpha_2, \ldots \alpha_n$ in this order. The precise definition of ψ_A will come later.

These waves must be summed up together with their phase. It is easy to see from standard scattering theory that a spherical wave decays as $\lambda|\text{distance}|^{-(d-1)/2}$ (the prefactor λ comes from the fact that the total amplitude of the scattering is scaled down by a factor of λ due to the coupling constant in front of the potential, see (5.10)). Since the amplitudes multiply, and since the typical mean free path is λ^{-2}, after n collisions the typical amplitude of ψ_A is

$$|\psi_A| \sim \left[\lambda\left(\lambda^2\right)^{\frac{d-1}{2}}\right]^n = \lambda^{dn}.$$

Thus if we try to sum up these waves in (6.1) with neglecting their phases, then

$$\sum_A |\psi_A| \sim |\Lambda|^n \lambda^{dn} = \left(\lambda^{-2d}\right)^n \lambda^{dn} = \lambda^{-dn} \to \infty.$$

However, if we can sum up these waves *assuming* that their phases are independent, i.e. they could be summed up as independent random variables, then it is the variance that is additive (exactly as in the central limit theorem in (2.1)):

$$|\psi_t|^2 = \Big|\sum_A \psi_A\Big|^2 \approx \sum_A |\psi_A|^2 \sim |\Lambda|^n \lambda^{2dn} = O(1). \tag{6.2}$$

Thus it is essential to extract a strong independence property from the phases. Since phases are determined by the random obstacles, there is some hope that at least most of the elementary waves are roughly independent. This will be our main technical achievement, although it will be formulated in a different language.

We remark that in the physics literature, the independence of phases is often postulated as an Ansatz under the name of "random phase approximation". Our result will mathematically justify this procedure.

6.2.2 Diffusive scale

Now we wish to go beyond the kinetic scale and describe a system with potentially infinitely many collisions. Thus we choose a longer time scale and we rescale the space appropriately. We obtain the following **diffusive scaling**:

$$t = \lambda^{-\kappa}\lambda^{-2}T, \qquad x = \lambda^{-\kappa/2}\lambda^{-2}T, \qquad n = \lambda^2 t = \lambda^{-\kappa}.$$

Notice that the time is rescaled by an additional factor $\lambda^{-\kappa}$ compared with the kinetic scaling, and space is rescaled by the square root of this additional factor. This represents the idea that the model scales diffusively with respect to the units of the kinetic scale. The total number of collisions is still $n = \lambda^2 t$, and now it tends to infinity.

If we try to translate this scaling into the elementary wave picture, then first we have to choose a box that is larger than the largest space scale, i.e. $L \geq \lambda^{-2-\kappa/2}$. The total number of elementary waves to sum up is

$$|\Lambda|^n = (\lambda^{-2-\frac{\kappa}{2}})^{dn} \sim \lambda^{-\lambda^{-\kappa}}$$

i.e. superexponentially large. Even with the assumption that only the variances have to be summable (see (6.2)), we still get a superexponentially divergent sum:

$$|\psi_t|^2 \approx \sum_A |\psi_A|^2 \sim |\Lambda|^n \lambda^{2nd} = \left[\lambda^{-2-\kappa/2}\right]^{nd} \lambda^{2nd} = \lambda^{-nd\kappa/2} = \lambda^{-(const)\lambda^{-\kappa}}$$

We will have to perform a **renormalization** to prevent this blow-up.

6.3 Kinetic scale: (linear) Boltzmann equation

The main result on the kinetic scale is the following theorem:

Theorem 6.1. *[**Boltzmann equation in the kinetic scaling limit**] Let the dimension be at least $d \geq 2$. Consider the random Schrödinger evolution on $\mathbb{R}^d$ or $\mathbb{Z}^d$*

$$i\partial_t \psi_t = H\psi_t, \qquad H = H_\lambda = -\frac{1}{2}\Delta + \lambda V(x),$$

where the potential is spatially uncorrelated (see more precise conditions below). Consider the kinetic rescaling

$$t = \lambda^{-2}\mathcal{T}, \qquad x = \lambda^{-2}\mathcal{X},$$

with setting $\varepsilon := \lambda^2$ to be the scale separation parameter in space.

Then the weak limit of the expectation of the rescaled Wigner transform of the solution ψ_t exists,

$$\mathbf{E}\, W^{\varepsilon}_{\psi_{\mathcal{T}/\varepsilon}}(\mathcal{X}, \mathcal{V}) \rightharpoonup F_{\mathcal{T}}(\mathcal{X}, \mathcal{V})$$

and F_T satisfies the linear Boltzmann equation,

$$\Big(\partial_{\mathcal{T}} + \nabla e(V) \cdot \nabla_{\mathcal{X}}\Big) F_{\mathcal{T}}(\mathcal{X}, V) = \int dU \sigma(U, V) \Big[F_{\mathcal{T}}(\mathcal{X}, U) - F_{\mathcal{T}}(\mathcal{X}, V)\Big] .$$

Here $e(V)$ is the dispersion relation of the free kinetic energy operator given by $e(V) = \frac{1}{2}V^2$ for the continuous model and by (5.3) for the discrete model. The collision kernel $\sigma(U, V)$ is explicitly computable and is model dependent (see below).

The velocities U, V are interpreted as incoming and outgoing velocities, respectively, see Fig. 8. The collision kernel always contains an *onshell condition*, i.e. a term $\delta(e(U) - e(V))$ guaranteeing energy conservation (see the examples below). Thus the right hand side of the linear Boltzmann equation is exactly the generator of the jump process on a fixed energy shell (in case of $e(V) = \frac{1}{2}V^2$ it is on the sphere) as defined in Section 2.4.

This theorem has been proven in various related models.

- [**Continuous model on** $\mathbb{R}^d$] The random potential is chosen to be a homogeneous Gaussian field, this means that the random variable $V(x)$ is Gaussian and it is stationary with respect to the space translations. We assume that $\mathbf{E}\, V(x) = 0$, then the distribution of $V(x)$ is uniquely determined by the two-point correlation function

$$R(x - y) := \mathbf{E} V(x) V(y),$$

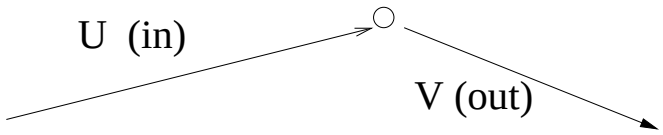

Figure 8: Incoming and outgoing momentum in a collision

and we assume that R is a Schwartz function. The dispersion relation of the free Laplacian on $\mathbb{R}^d$ is

$$e(p) := \frac{1}{2}p^2,$$

and the Boltzmann collision kernel is computed from the microscopic data as

$$\sigma(U, V) = \delta(e(U) - e(V))|\widehat{R}(U - V)|, \tag{6.3}$$

This result was first obtained by Spohn [45] for short macroscopic time, $T \leq T_0$, and later extended to arbitrary times in a joint work with Yau [21] with a different method. We assume that the dimension $d \geq 2$, with a special non-degeneracy condition on the initial data in case of $d = 2$ forbidding concentration of mass at zero momentum. The $d = 1$ case is special; due to strong interference effects no Markovian limiting dynamics is expected.

- [**Discrete model + Non Gaussian**] The method of [21] was extended by Chen in two different directions [9]. He considered the discrete model, i.e. the original Anderson model and he allowed non-Gaussian randomness as well. In the case of Laplacian on $\mathbb{Z}^d$ the dispersion relation is more complicated (5.3), but the collision kernel is simpler;

$$\sigma(U, V) = \delta(e(U) - e(V)), \tag{6.4}$$

i.e. it is simply the uniform measure on the energy surface given by the energy of the incoming velocity.

- [**Phonon model**] In this model the random potential is replaced by a heat bath of non-interacting bosonic particles that interact only with the single electron. *Formally*, this model leads to an electron in a time dependent random potential, but the system is still Hamiltonian. This model is the quantum analogue of Einstein's picture. The precise formulation of the result is a bit more complicated, since the phonon bath has to be mathematically introduced, so we will not do it here. The interested reader can consult with [18]. More recently, De Roeck and Fröhlich [10] have proved diffusion for any times in $d \geq 4$ dimensions with an additional strong spin coupling that enhances the loss of memory.

- [**Cubic Schrödinger equation with random initial data**] A nonlinear version of the Boltzmann equation was proven in [39] where the Schrödinger equation contained a weak cubic nonlinearity and the initial data was drawn from near thermal equilibrium.

- [**Wave propagation in random medium**] It is well known that in a system of harmonically coupled oscillators the wave propagate ballistically. If the masses of the oscillators deviate randomly from a constant value, the wave propagation can change. The effect is similar to the disturbances in the free Schrödinger propagation; the ballistic transport is a coherence effect that can be destroyed by small perturbations. This model, following the technique of [21], has been worked out in [38].

- [**Low density models**] There is another way to reduce the effects of possible recollisions apart from introducing a small coupling constant: one can consider a low density of obstacles. Let V_0 be a radially symmetric Schwartz function with a sufficiently small weighted Sobolev norm:

$$\|\langle x\rangle^{N(d)}\langle\nabla\rangle^{N(d)}V_0\|_\infty \qquad \text{is sufficiently small for some } N(d) \text{ sufficiently large.}$$

  The Hamiltonian of the model is given by

$$H = -\frac{1}{2}\Delta_x + \sum_{\alpha=1}^{M} V_0(x - x_\alpha) \qquad \text{in a box } [-L, L]^d, \quad L \gg \varepsilon^{-1}$$

  acting on $\mathbb{R}^d$, $d \geq 3$, where $\{x_\alpha\}_{\alpha=1,\dots M}$ denotes a collection of random i.i.d. points. with density

$$\varrho := \frac{M}{L^d} \to 0.$$

  The kinetic scaling corresponds to choosing

$$x \sim \varepsilon^{-1}, \ t \sim \varepsilon^{-1}, \qquad \varepsilon = \varrho,$$

  and then letting $\varepsilon \to 0$. In this case convergence to the linear Boltzmann equation in the spirit of Theorem 6.1 was proven in [20], [17] (although these papers consider the $d = 3$ case, the extension to higher dimensions is straightforward). The dispersion relation is $e(p) = \frac{1}{2}p^2$ and the collision kernel is

$$\sigma(U, V) = |T(U, V)|^2\delta(U^2 - V^2)$$

  where $T(U, V)$ is the quantum scattering cross section for $-\frac{1}{2}\Delta + V_0$.

It is amusing the compare the weak-coupling and the low-density sceneries; they indeed describe different physical models (Fig. 9). Although in both cases the result is the linear Boltzmann equation, the microscopic properties of the model influence the collision kernel. In particular, in the low density model the full single-bump scattering information is needed (while the function $\widehat{R}$ (6.3) in the weak coupling model can be viewed as the Born approximation of the full scattering cross section). The scaling is chosen such that the typical number of collisions remains finite in the scaling limit. Note that in both models the wavelength is comparable with the spatial variation of the potential; both live on the microscopic scale. Thus neither model is semiclassical.

We make one more interesting remark. When compared with the corresponding classical Hamiltonian, the low density model yields the (linear) Boltzmann equation both in classical

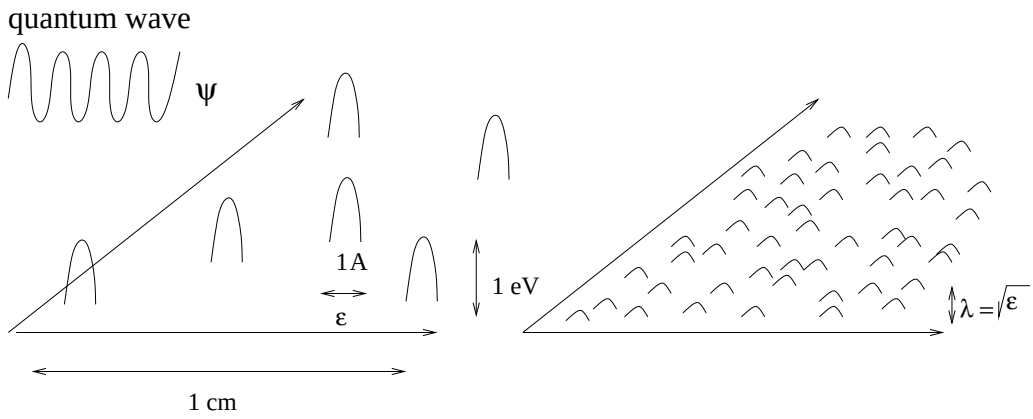

Figure 9: Low density and weak coupling sceneries

and quantum mechanics (although the collision kernels are different). The weak coupling model yields Boltzmann equation when starting from quantum mechanics, however it yields a random walk on the energy shell (for the limiting velocity process v_t) when starting from classical mechanics [33] and [35]. For further references, see [21].

Fig. 10 depicts schematically the microscopic and macroscopic scales in Theorem 6.1. The right side of the picture is under a microscope which is "sees" scales comparable with the wavelength (Angstrom scale). Only one obstacle and only one outgoing wave are pictured. In reality the outgoing wave goes in all directions. On this scale the Schrödinger equation holds:

$$i\partial_t \psi_t(x) = \Big[-\Delta_x + \lambda V(x) \Big] \psi_t(x).$$

On the left side we zoomed our "microscope" out. Only a few obstacles are pictured that are touched by a single elementary wave function. All other trajectories are also possible. The final claim is that on this scale the dynamics can be described by the Boltzmann equation

$$\Big(\partial_{\mathcal{T}} + \nabla e(V) \cdot \nabla_{\mathcal{X}} \Big) F_{\mathcal{T}}(\mathcal{X}, V) = \int \mathrm{d}U \sigma(U, V) \Big[F_{\mathcal{T}}(\mathcal{X}, U) - F_{\mathcal{T}}(\mathcal{X}, V) \Big].$$

Actually the obstacles (black dots) on the left picture are only fictitious: recall that there are no physical obstacles behind the Boltzmann equation. It represents a process, where physical obstacles are replaced by a stochastic dynamics: the particle has an exponential clock and it randomly decides to change its trajectory i.e. to behave as if there were an obstacle (recall Section 2.4). It is crucial to make this small distinction because it is exactly the fictitious obstacles are the signatures that all recollisions have been eliminated hence this guarantees the Markov property for the limit dynamics.

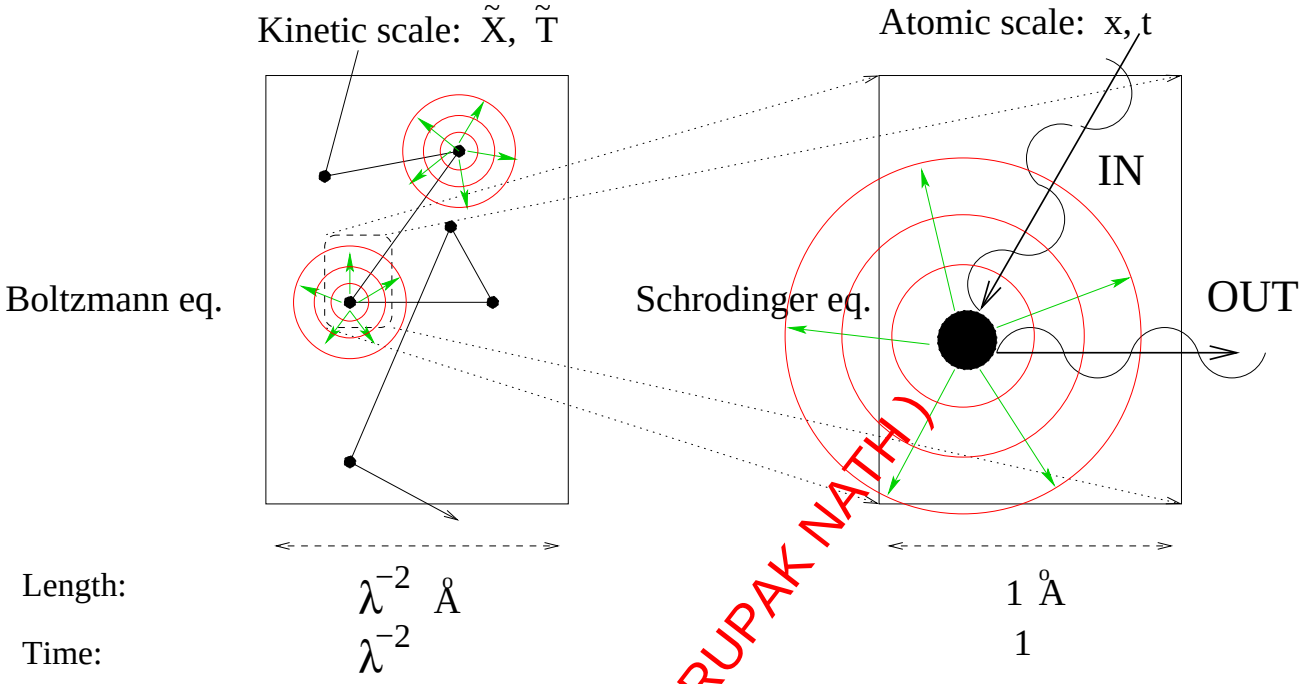

Figure 10: Macroscopic and microscopic scales

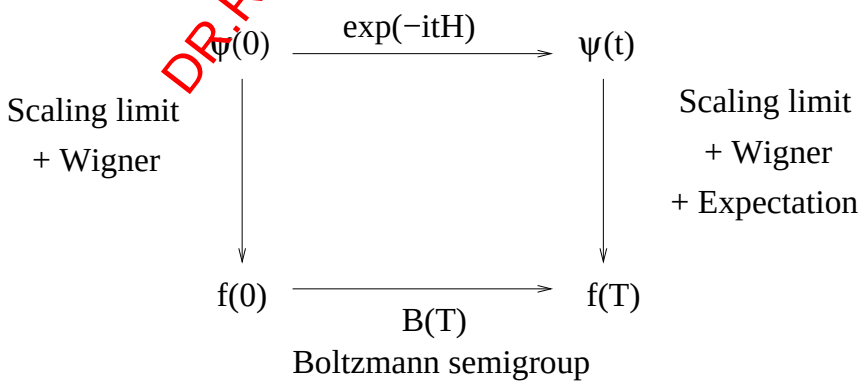

Figure 11: Boltzmann semigroup models Schrödinger evolution

The main message of the result is that the long time ($t = T\varepsilon^{-1}$) Schrödinger evolution can modelled by a finite time (T) Boltzmann evolution on the macroscopic scale. Of course the detailed short scale information is lost, which explains irreversibility. The effective limiting equation is classical, but quantum features are retained in the collision kernel.

Note that this approximation is not just conceptually interesting but it also gives rise to an enormous computational simplification. Solving the Schrödinger equation directly for a long time in an complicated environment is a computationally impossible task. In contrast, the Boltzmann equation needs to be solved on $O(1)$ times and there is no complicated environment. This is expressed schematically in the "commutative diagram" Fig. 11. The top horizontal line is the true (Schrödinger) evolution. The theorem says that one can "translate" the quantum initial data by Wigner transform and scaling limit into a phase space density $f(0) = f_0(X, V)$ and one can use the much simpler (and shorter time) Boltzmann evolution to f_0 to obtain the Wigner transform (after expectation and scaling limit) of the true evolution ψ_t.

6.4 Diffusive scale: Heat equation

We consider the same model with a weakly coupled random potential as before. For definiteness, we work on the lattice $\mathbb{Z}^d$, $d \geq 3$, the Hamiltonian is given by

$$H = -\Delta + \lambda V,$$

where the discrete Laplacian is given by (5.2) with dispersion relation (5.3) and where the random potential is

$$V(x) = \sum_{\alpha \in \mathbb{Z}^d} v_\alpha \delta(x - \alpha).$$

The random variables v_α are i.i.d. with moments $m_k = \mathbf{E}\, v_\alpha^k$ satisfying

$$m_1 = m_3 = m_5 = 0, \quad m_{2d} < \infty.$$

Let

$$\psi_t := e^{-itH}\psi_0,$$

and we rescale the Wigner transform as before:

$$W_{\psi_t} \to W^\varepsilon_{\psi_t} = \varepsilon^{-d} W_{\psi_t}(X/\varepsilon, v).$$

We remark that in case the discrete lattice $\mathbb{Z}^d$ the definition of the Wigner function $W_\psi(x, v)$ is given by (4.2). Our main result is the following

Theorem 6.2. [**Quantum diffusion in the discrete case**] *For any dimension $d \geq 3$ there exists $\kappa_0(d) > 0$ [in $d = 3$ one can choose $\kappa_0(3) = \frac{1}{10000}$], such that for any $\kappa \leq \kappa_0(d)$ and any $\psi_0 \in L^2(\mathbb{Z}^d)$ the following holds. In the diffusive scaling*

$$t = \lambda^{-\kappa}\lambda^{-2}T, \qquad x = \lambda^{-\kappa/2}\lambda^{-2}X, \qquad \varepsilon = \lambda^{-\kappa/2-2},$$

we have that

$$\int_{\{e(v)=e\}} \mathbf{E}\, W^\varepsilon_{\psi_t}(X, v)\, \mathrm{d}v \rightharpoonup f_T(X, e) \qquad \text{weakly as } \lambda \to 0, \tag{6.5}$$

and for almost all $e > 0$ the limiting function satisfies the heat equation

$$\partial_T f_T(X,e) = \nabla_X \cdot D(e) \nabla_X f_T(X,e) \tag{6.6}$$

with a diffusion matrix given by

$$D_{ij}(e) = \Big\langle \nabla e(v) \otimes \nabla e(v) \Big\rangle_e, \qquad \langle f(v) \rangle_e = \textit{Average of } f \textit{ on } \{v \,:\, e(v) = e\} \tag{6.7}$$

and with initial state

$$f_0(X,e) := \delta(X) \int \delta(e(v) - e) |\widehat{\psi}_0(v)|^2 \mathrm{d}v.$$

The weak convergence in (6.10) *means that for any Schwartz function $J(x,v)$ on $\mathbb{R}^d \times \mathbb{R}^d$ we have*

$$\lim_{\varepsilon \to 0} \int_{\left(\frac{\varepsilon}{2}\mathbb{Z}\right)^d} \mathrm{d}X \int \mathrm{d}v J(X,v) \mathbf{E}\, W^{\varepsilon}_{\psi(\lambda^{-\kappa-2}T)}(X,v) = \int_{\mathbb{R}^d} \mathrm{d}X \int \mathrm{d}v J(X,v) f_T(X, e(v))$$

uniformly in T on $[0, T_0]$, where T_0 is an arbitrary fixed number.

This result is proven in [25]. The related continuous model is treated in [23, 24]. A concise summary of the ideas can be found in the expository paper [22]. Note again that weak limit means that only macroscopic observables can be controlled. Moreover, the theorem does not keep track of the velocity any more, since on the largest diffusive scale the velocity of the particle is not well defined (similarly as the Brownian motion has no derivative). Thus we have to integrate out the velocity variable on a fixed energy shell (the energy remains macroscopic observable). This justifies the additional velocity integration in (6.5).

The diffusion matrix is the velocity autocorrelation matrix (computed in equilibrium) obtained from the Boltzmann equation

$$D_{Boltz}(e) = \int_0^\infty \mathcal{E}_e \left[\nabla e(v(0)) \otimes \nabla e(v(t)) \right], \mathrm{d}t \tag{6.8}$$

similarly to Theorem 2.3. Here $\mathcal{E}_e$ denotes the expectation value of the Markov process $v(t)$ described by the linear Boltzmann equation as its generator if the initial velocity $v(0)$ is distributed according to the equilibrium measure of the jump process with generator (2.10) on the energy surface $\{e(v) = e\}$ with the Boltzmann collision kernel (6.4). Since this kernel is the uniform measure on the energy surface, the Boltzmann velocity process has no memory and thus it is possible to compute the velocity autocorrelation just by averaging with respect to the uniform measure on $\{e(v) = e\}$:

$$D_{Boltz}(e) = \int_0^\infty \mathcal{E}_e \left[\nabla e(v(0)) \otimes \nabla e(v(t)) \right] \mathrm{d}t = \Big\langle \nabla e(v) \otimes \nabla e(v) \Big\rangle_e. \tag{6.9}$$

In particular, Theorem 6.2 states that the diffusion matrix $D(e)$ obtained from the quantum evolution coincides with the velocity autocorrelation matrix of the linear Boltzmann equation.

For completeness, we state the result also for the continuum model.

Theorem 6.3. [Quantum diffusion in the continuum case] *Let the dimension be at least $d \geq 3$ and $\psi_0 \in L^2(\mathbb{R}^d)$. Consider the diffusive scaling*

$$t = \lambda^{-\kappa}\,\lambda^{-2}T, \qquad x = \lambda^{-\kappa/2}\,\lambda^{-2}X, \qquad \varepsilon = \lambda^{-\kappa/2-2}.$$

Let

$$\psi_t = e^{itH}\psi_0, \qquad H = -\frac{1}{2}\Delta + \lambda V,$$

where the random potential is given by (5.5) with a single site profile B that is a spherically symmetric Schwartz function with 0 in the support of its Fourier transform, $0 \in \operatorname{supp}\widehat{B}$. For $d \geq 3$ there exists $\kappa_0(d) > 0$ (in $d = 3$ one can choose $\kappa_0(3) = 1/370$) such that for any $\kappa < \kappa_0(d)$ we have

$$\int_{\{e(v)=e\}} \mathbf{E}\, W^{\varepsilon}_{\psi_t}(X,v)\,\mathrm{d}v \rightharpoonup f_T(X,e) \qquad \text{(weakly as } \lambda \to 0\text{)}, \tag{6.10}$$

and for almost all energies $e > 0$ the limiting function satisfies the heat equation

$$\partial_T f_T(X,e) = D_e \Delta_X f_T(X,e)$$

with initial state

$$f_0(X,e) := \delta(X) \int \delta(e(v) - e)|\widehat{\psi}_0(v)|^2 \mathrm{d}v.$$

The diffusion coefficient is given by

$$D_e := \frac{1}{d}\int_0^\infty \mathcal{E}_e\left[v(0)\cdot v(t)\right]\,\mathrm{d}t, \tag{6.11}$$

where $\mathcal{E}_e$ is the expectation value for the random jump process on the energy surface $\{e(v) = e\}$ with generator

$$\sigma(v,v') = |\widehat{B}(v - v')|^2 \delta(e(v) - e(v')).$$

The condition $0 \in \operatorname{supp}\widehat{B}$ is not essential for the proof, but the statement of the theorem needs to be modified if the support is separated away from zero. In this case, the low momentum component of the initial wave function moves ballistically since the diameter of the energy surface is smaller than the minimal range of $\widehat{B}$. The rest of the wave function still moves diffusively.

Fig. 12 shows schematically the three scales we discussed. Notice that going from the Boltzmann scale to the diffusive scale is essentially the same as going from random walk to Brownian motion in Wiener's construction (Theorems 2.3 and 2.7). However it is misleading to try to prove Theorem 6.2 by a two-step limiting argument, first using the kinetic limit (Theorem 6.1) then combining it with Wiener's argument. There is only one limiting parameter in the problem, one cannot take first the kinetic limit, then the additional $\lambda^{-\kappa} \to \infty$ limit modelling the long kinetic time scale. The correct procedure is to consider the Schrödinger evolution and run it up to the diffusive time scale. In this way, the intermediate Boltzmann scale can give only a hint what to expect but it cannot be used for proof.

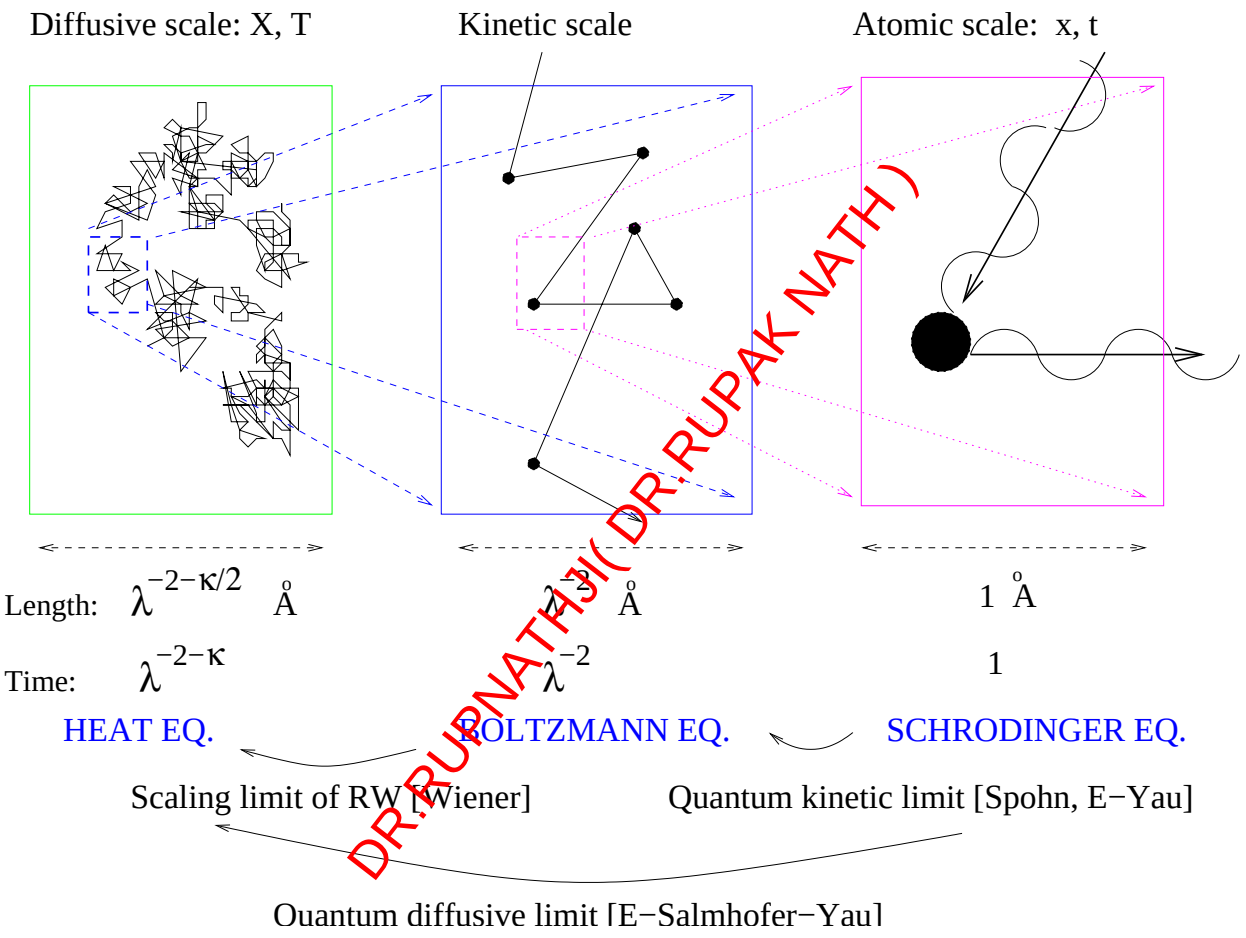

Figure 12: Three scales: diffusive, kinetic and atomic

In fact, this is not only a mathematical subtlety. Correlations that are small on the kinetic scale, and were neglected in the first limit, can become relevant on longer time scales. Theorem 6.2 shows that it is not the case; at least up to times $t \sim \lambda^{-2-\kappa}$ the correlations are not strong enough to destroy the diffusive picture coming from the long time limit of the Boltzmann equation (in particular the diffusion coefficient can be correctly computed from the Boltzmann equation). However, this should not be taken for granted, and in fact it is not expected to hold in $d = 2$ dimensions for exponentially long times, $t \sim \exp(\lambda^{-1})$. Theorem 6.1 on the kinetic limit is valid in $d = 2$ as well and Wiener's argument is dimension independent. On the other hand if the diffusive evolution were true up to any time scales, then in particular the state would be delocalized, in contrast to the conjecture that the random Schrödinger operator in $d = 2$ dimensions is always localized.

7 Feynman graphs (without repetition)

7.1 Derivation of Feynman graphs

Feynman graphs are extremely powerful graphical representations of perturbation expansions. They have been primarily used in many-body theories (they were invented for QED), but they are very convenient for our problem as well to organize perturbation expansion. They have been used by physicist to study random Schrödinger operators, a good overview from the physics point of view is [46].

We will derive the Feynman graphs we need for this presentation. Recall that

$$H = -\Delta + \lambda V, \qquad V = \sum_{\alpha \in \mathbb{Z}^d} V_\alpha, \quad \mathbf{E} V_\alpha = 0, \tag{7.1}$$

where V_α is a single-bump potential around $\alpha \in \mathbb{Z}^d$, for example

$$V_\alpha(x) = v_\alpha \delta(x - \alpha).$$

We consider the potential as a small perturbation of $-\Delta$ and we expand the unitary evolution by the identity (Duhamel formula)

$$\psi_t = e^{-itH}\psi_0 = e^{it\Delta}\psi_0 - i\lambda \int_0^t e^{-i(t-s)H} V e^{is\Delta}\psi_0 \, \mathrm{d}s. \tag{7.2}$$

This identity can be seen by differentiating $U(t) := e^{-itH}e^{itH_0}$ where $H = H_0 + \lambda V$, $H_0 = -\Delta$:

$$\frac{\mathrm{d}U}{\mathrm{d}t} = e^{-itH}(-iH + iH_0)e^{itH_0} = e^{-itH}(-i\lambda V)e^{itH_0},$$

thus we can integrate back from 0 to t

$$U(t) = I + \int_0^t \mathrm{d}s \; e^{-i(t-s)H}(-i\lambda V)e^{i(t-s)H_0}$$

and multiply by e^{-itH_0} from the right.

We can iterate the expansion (7.2) by repeating the same identity for $e^{-i(t-s)H}$. For example, the next step is

$$e^{-itH}\psi_0 = e^{it\Delta}\psi_0 - i\lambda\int_0^t e^{i(t-s)\Delta}Ve^{is\Delta}\psi_0\,\mathrm{d}s + \lambda^2\int_0^t \mathrm{d}s_1\int_0^{s_1}\mathrm{d}s_2\; e^{-i(t-s_1-s_2)H}Ve^{is_2\Delta}Ve^{is_1\Delta}\psi_0$$

etc. We see that at each step we obtain a new *fully expanded* term (second term), characterized by the absence of the unitary evolution of H; only free evolutions $e^{-is\Delta}$ appear in these terms. There is always one last term that contains the full evolution and that will be estimated trivially after sufficient number of expansions. More precisely, we can expand

$$\psi_t = e^{-itH}\psi_0 = \sum_{n=0}^{N-1}\psi^{(n)}(t) + \Psi_N(t), \tag{7.3}$$

where

$$\psi^{(n)}(t) := (-i\lambda)^n\int_{\mathbb{R}_+^{n+1}}\mathrm{d}s_0\mathrm{d}s_1\ldots\mathrm{d}s_n\;\delta\Big(t-\sum_{j=0}^n s_j\Big)e^{is_0\Delta}Ve^{is_1\Delta}V\ldots Ve^{is_n\Delta}\psi_0$$

and

$$\Psi_N(t) := (-i\lambda)\int_0^t \mathrm{d}s\; e^{-i(t-s)H}V\psi^{(N-1)}(s). \tag{7.4}$$

Recalling that each V is a big summation (7.1), we arrive at the following *Duhamel formula*

$$\psi_t = \sum_{n=0}^{N-1}\sum_A \psi_A + \text{full evolution term}, \tag{7.5}$$

where the summation is over collision histories:

$$A = (\alpha_1, \alpha_2, \cdots, \alpha_n), \qquad \alpha_j \in \mathbb{Z}^d,$$

and each elementary wave function is given by

$$\psi_A := \psi_A(t) = (-i\lambda)^n\int \mathrm{d}\mu_{n,t}(\underline{s})\; e^{is_0\Delta}\; V_{\alpha_1}\cdots V_{\alpha_n}e^{is_n\Delta}\;\psi_0. \tag{7.6}$$

Here, for simplicity, we introduced the notation

$$\int \mathrm{d}\mu_{n,t}(\underline{s}) := \int_{\mathbb{R}_+^{n+1}}\delta\Big(t-\sum_{j=0}^n s_j\Big)\mathrm{d}s_0\mathrm{d}s_1\ldots\mathrm{d}s_n \tag{7.7}$$

for the integration over the simplex $s_0 + s_1 + \ldots + s_n = t$, $s_j \geq 0$.

If the cardinality of A is $n = |A|$, then we say that the elementary wave function is of order n and their sum for a fixed n is denoted by

$$\psi^{(n)}(t) := \sum_{A\,:\,|A|=n}\psi_A(t).$$

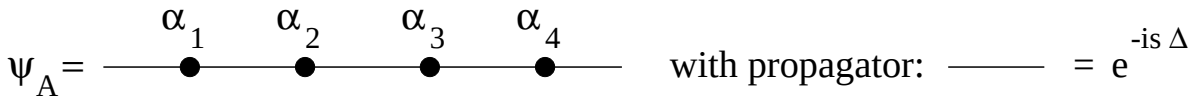

Figure 13: Representation of a wave function with collision history A

Note that the full evolution term Ψ_N has a similar structure, but the leftmost unitary evolution $e^{is_0\Delta}$ is replaced with $e^{-is_0 H}$.

We remark that not every elementary wave function has to be expanded up to the same order. The expansion can be adjusted at every step, i.e. depending on the past collision history, encoded in the sequence A, we can decice if an elementary wave function ψ_A should be expanded further or we stop the expansion. This will be essential to organize our expansion without overexpanding: once the collision history A indicates that ψ_A is already small (e.g. it contains recollisions), we will not expand it further.

Each ψ_A can be symbolically represented as shown on Fig. 13, where lines represent the free propagators and bullets represent the potential terms (collisions). Note the times between collisions is not recorded in the graphical picture since they are integrated out.

We need to compute expectation values of quadratic functionals of ψ like the Wigner transform. For simplicity we will show how the L^2-norm can be treated, the Wigner transform is very similar (see next Section).

Suppose for the moment that there is no repetition in A, i.e. all α_j's are distinct. To compute $\mathbf{E}\|\psi\|^2 = \mathbf{E}\,\overline{\psi}\,\psi$, we need to compute

$$\sum_{n,n'}\sum_{A\,:|A|=n}\sum_{B:|B|=n'} \mathbf{E}\,\langle\psi_B\,,\psi_A\rangle.$$

This is zero unless B is a permutation of A, because individual potentials have zero expectation. Higher order moments of the same potential V_α are excluded since we assumed that A and B have no repetition. Thus the only nonzero contributions come from second moments, i.e. from pairings of each element of A with an element of B, in particular, $|A| = |B|$ is forced, i.e. $n = n'$.

In conclusion, with the tacit assumption that there are no repetitions, we can express $\mathbf{E}\|\psi\|^2$ as a summation over the order n and over permutations between the sets A and B with $|A| = |B| = n$:

$$\sum_{n,n'}\sum_{A\,:|A|=n}^{*}\sum_{B:|B|=n'}^{*} \mathbf{E}\,\langle\psi_B\,,\psi_A\rangle = \sum_{n}\sum_{\pi\in S_n}\sum_{A\,:|A|=n}^{*} \mathbf{E}\,\langle\psi_{\pi(A)}\,,\psi_A\rangle =: \sum_{n}\sum_{\pi\in S_n}\mathrm{Val}(\pi), \qquad (7.8)$$

where S_n is set of permutations on n elements and upper star denotes summation over non-repetitive sequences. $\mathrm{Val}(\pi)$ is defined by the last formula by summing up all sequences A; this summation will be computed explicitly and it will give rise to delta functions among momenta.

Drawing the two horizontal lines that represent ψ_A and ψ_B parallel and connecting the paired bullets, we obtain Fig. 14, where the first summation is over the order n, the second summation is over all permutations $\pi \in S_n$. a graph, called the *Feynman graph of* π. We will give an explicit formula for the value of each Feynman graph, here denoted by $\mathrm{Val}(\pi)$.

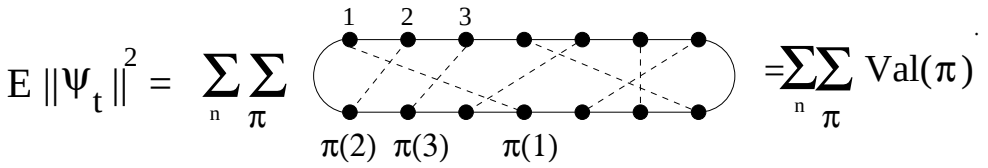

Figure 14: Representation of the L^2-norm as sum of Feynman graphs

7.2 L^2-norm vs. Wigner transform

Eventually we need to express

$$\mathbf{E}\langle J, W_\psi\rangle := \int J(x,v)\mathbf{E}\, W_\psi(x,v)\mathrm{d}x\mathrm{d}v = \int \widehat{J}(\xi,v)\mathbf{E}\,\widehat{W}_\psi(\xi,v)\mathrm{d}\xi\mathrm{d}v = \mathbf{E}\langle \widehat{J}, \widehat{W}_\psi\rangle$$

(recall that hat means Fourier transform only in the x variable). However, we have

Lemma 7.1. *For any Schwarz function J, the quadratic functional $\psi \mapsto \mathbf{E}\langle \widehat{J}, \widehat{W}_\psi\rangle$ is continuous in the L^2-norm.*

Proof. Let $\psi, \phi \in L^2$ and set

$$\Omega := \mathbf{E}\Big(\langle \widehat{J}, \widehat{W}_\psi\rangle - \langle \widehat{J}, \widehat{W}_\phi\rangle\Big) = \mathbf{E}\int \mathrm{d}\xi\mathrm{d}v \widehat{J}(\xi,v)\Big[\widehat{W}_\psi(\xi,v) - \widehat{W}_\phi(\xi,v)\Big].$$

Let $v_\pm = v \pm \frac{\xi}{2}$ and write

$$\Big[\,\cdot\,\Big] = \overline{\widehat{\psi}(v_-)}\big(\widehat{\psi}(v_+) - \widehat{\phi}(v_+)\big) + \big(\overline{\widehat{\psi}(v_-)} - \overline{\widehat{\phi}(v_-)}\big)\widehat{\phi}(v_+).$$

Thus

$$\begin{aligned}|\Omega| &\le \int \mathrm{d}\xi\mathrm{d}v|\widehat{J}(\xi,v)|\,\mathbf{E}\Big|\Big[\,\cdot\,\Big]\Big| \le \int \mathrm{d}\xi \sup_v|\widehat{J}(\xi,v)|\int \mathrm{d}v\;\mathbf{E}\Big|\big[\,\cdot\,\big]\Big| \\ &\le \int \mathrm{d}\xi \sup_v|\widehat{J}(\xi,v)|\int \mathrm{d}v\;\mathbf{E}\Big[|\widehat{\psi}(v_-)|\cdot|\widehat{\psi}(v_+) - \widehat{\phi}(v_+)| + |\widehat{\psi}(v_-) - \overline{\widehat{\phi}(v_-)}|\cdot|\widehat{\phi}(v_+)|\Big].\end{aligned}$$

By Schwarz inequality,

$$\begin{aligned}\int \mathrm{d}v\mathbf{E}\Big(|\widehat{\psi}(v_-)|\cdot|\widehat{\psi}(v_+) - \widehat{\phi}(v_+)|\Big) &\le \Big(\int \mathrm{d}v\mathbf{E}|\widehat{\psi}(v_-)|^2\Big)^{1/2}\Big(\int \mathrm{d}v\mathbf{E}|\widehat{\psi}(v_+) - \widehat{\phi}(v_+)|^2\Big)^{1/2} \\ &= \Big(\int \mathrm{d}v\mathbf{E}|\widehat{\psi}(v)|^2\Big)^{1/2}\Big(\int \mathrm{d}v\mathbf{E}|\widehat{\psi}(v) - \widehat{\phi}(v)|^2\Big)^{1/2} \\ &= \Big(\mathbf{E}\|\widehat{\psi}\|^2\Big)^{1/2}\Big(\mathbf{E}\|\widehat{\psi} - \widehat{\phi}\|^2\Big)^{1/2}. \qquad (7.9)\end{aligned}$$

Thus

$$|\Omega| \le \Big(\int \mathrm{d}\xi \sup_v|\widehat{J}(\xi,v)|\Big)\Big(\sqrt{\mathbf{E}\|\psi\|^2} + \sqrt{\mathbf{E}\|\phi\|^2}\Big)\Big(\mathbf{E}\|\widehat{\psi} - \widehat{\phi}\|^2\Big)^{1/2},$$

which completes the proof of the Lemma. □

This lemma, in particular, guarantees that it is sufficient to consider nice initial data, i.e. we can assume that ψ_0 is a Schwarz function. More importantly, this lemma guarantees that it is sufficient to control the Feynman diagrammatic expansion of the L^2-norm of ψ; the same expansion for the Wigner transform will be automatically controlled. Thus all estimates of the error terms can be done on the level of the L^2-norm; the more complicated arguments appearing in the Wigner transform are necessary only for the explicit computation of the main term.

7.3 Stopping the expansion

We will never work with infinite expansions, see (7.3), but then we will have to control the last, fully expanded term. This will be done by using the unitarity of the full evolution $e^{-i(t-s)H}$ in the definition of Ψ_N, see (7.4). More precisely, we have:

Lemma 7.2 (Unitarity bound)**.** *Assume that $\|\psi_0\| = 1$. According to (7.3), we write $\psi(t) = \Phi_N(t) + \Psi_N(t)$ with $\Phi_N = \sum_{n=0}^{N-1} \psi^{(n)}$ containing the fully expanded terms. Then*

$$\Big|\mathbf{E}\Big[\langle J, W_{\psi(t)}\rangle - \langle J, W_{\Phi_N(t)}\rangle\Big]\Big| \le \Big(1 + \sqrt{\mathbf{E}\|\Phi_N\|^2}\Big)t\Big[\sup_{0\le s\le t} \mathbf{E}\|\lambda V\psi^{(N-1)}(s)\|^2\Big]^{1/2}.$$

In other words, the difference between the true wave function $\psi(t)$ and its N-th order approximation Φ_N *can be expressed in terms of fully expanded quantities*, but we have to pay an additional price t. This additional t factor will be compensated by a sufficiently long expansion, i.e. by choosing N large enough so that $\psi^{(N-1)}(s)$ is small. In practice we will combine this procedure by stopping the expansion for each elementary wave function ψ_A separately, depending on the collision history. Once ψ_A is sufficiently small to compensate for the t factor lost in the unitarity estimate, we can stop its expansion.

Proof. We apply Lemma 7.1 with $\psi = \psi(t)$ and $\phi = \Phi_N$, so that $\psi - \phi = \Psi_N$. Then we have, for any N, that

$$\Big|\mathbf{E}\Big[\langle J, W_{\psi(t)}\rangle - \langle J, W_{\Phi_N(t)}\rangle\Big] \le \Big(1 + \sqrt{\mathbf{E}\|\Phi_N\|^2}\Big)\Big(\mathbf{E}\|\Psi_N\|^2\Big)^{1/2},$$

since $\|\psi_t\| = \|\psi_0\| = 1$. Then using the unitarity, i.e. that

$$\Big\|e^{-i(t-s)H}\lambda V\psi^{(N-1)}(s)\Big\| = \Big\|\lambda V\psi^{(N-1)}(s)\Big\|,$$

we can estimate

$$\begin{aligned}
\mathbf{E}\|\Psi_N\|^2 &= \mathbf{E}\Big\|\int_0^t \mathrm{d}s e^{-i(t-s)H}\lambda V\psi^{(N-1)}(s)\Big\|^2 \\
&\le \mathbf{E}\Big(\int_0^t \mathrm{d}s\Big\|\lambda V\psi^{(N-1)}(s)\Big\|\Big)^2 \\
&\le t\int_0^t \mathrm{d}s\ \mathbf{E}\Big\|\lambda V\psi^{(N-1)}(s)\Big\|^2, \qquad (7.10)
\end{aligned}$$

which proves Lemma 7.2. □

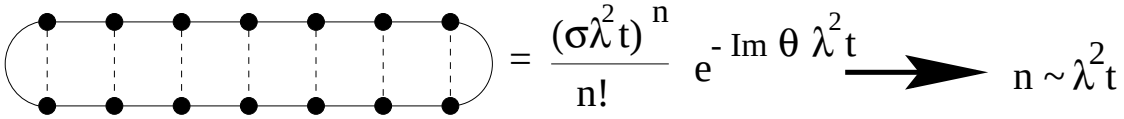

Figure 15: Contribution of the ladder graph

7.4 Outline of the proof of the Boltzmann equation with Feynman graphs

The main observation is that among the $n!$ possible permutations of order n, only one, the identity permutation contributes to the limiting Boltzmann equation. Actually it has to be like that, since this is the only graph whose collision history can be interpreted classically, since here $A = B$ as sets, so the two wavefunctions ψ_A and ψ_B visit the same obstacles in the same order. If there were a discrepancy between the order in which the obstacles in A and B are visited, then no classical collision history could be assigned to that pairing.

The Feynman graph of the identity permutation is called *the ladder graph*, see Fig. 15. Its value can be computed explicitly and it resembles to the n-th term in the Taylor series of the exponential function.

The ladder with n pairings corresponds to a classical process with n collisions. If the collision rate is $\sigma\lambda^2$, then the classical probability of n collisions is given by the (unnormalized) Poisson distribution formula

$$\text{Prob } (n \text{ collisions}) = \frac{(\sigma\lambda^2 t)^n}{n!}.$$

This is clearly summable for the kinetic scale $t \sim \lambda^{-2}$, and the typical number of collisions is the value n giving the largest term, i.e. $n \sim \lambda^2 t$. The exponential damping factor $e^{-\text{Im}(\theta)\lambda^2 t}$ in Fig. 15 comes from renormalization (see later in Section 8.4 where the constants σ and θ will also be defined).

We note that from the Duhamel formula leading to the Feynman graphs, it is not at all clear that the value of the ladder graphs is of order one (in the limiting parameter λ). According to the Duhamel representation, they are highly oscillatory integrals of the form (written in momentum space)

$$\text{Val}(id_n) \sim \lambda^{2n} \Big(\int_0^t e^{is_1 p_1^2} \int_0^{s_1} e^{is_2 p_2^2} \dots \int_0^{s_n} \Big) \Big(\int_0^t e^{-is'_1 p_1^2} \int_0^{s'_1} e^{-is'_2 p_2^2} \dots \int_0^{s'_n} \Big). \tag{7.11}$$

Here id_n is the identity permutation in S_n which generates the ladder. If one estimates these integrals naively, neglecting all oscillations, one obtains

$$\text{Val}(id_n) \leq \frac{\lambda^{2n} t^{2n}}{(n!)^2},$$

which is completely wrong, since it does not respect the power counting that $\lambda^2 t$ should be the relevant parameter.

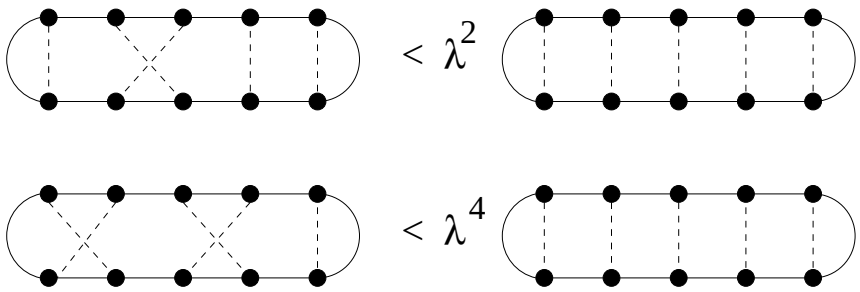

Figure 16: Comparing the value of various crossing graphs

Assuming that the ladder can be computed precisely, with all oscillations taken into account, one still has to deal with the remaining $n!-1$ non-ladder graphs. Their explicit formulas will reveal that their values are small, but there are many of them, so it is not clear at all if they are still summable to something negligible.

It is instructive to compare the value of the non-ladder graphs with the ladder graph because that is of order one (see Section 8.6 for a detailed computation). Fig. 16 contains schematic estimates of simple crossing graphs that can be checked by direct computation (once the explicit formula for $\mathrm{Val}(\pi)$ is established in the next section). It indicates that the value of a non-ladder graph is a positive power of λ that is related with the combinatorial complexity of the permutation (e.g. the number of "crosses").

Accepting for the moment that all non-ladder graphs are at least by an order λ^2 smaller than the ladder graphs, we obtain the following majorizing estimate for the summation of the terms in the Duhamel expansion:

$$\sum_n \sum_\pi |\mathrm{Val}(\pi)| \leq \sum_n \frac{(\sigma\lambda^2 t)^n}{n!} \Big(\underbrace{1}_{A=B} + \underbrace{(n!-1)(\text{small})}_{A\neq B} \Big).$$

This series clearly converges for short kinetic time ($\lambda^2 t = T \leq T_0$) but not for longer times, since the two $n!$'s cancel each other and the remaining series is geometric. This was the basic idea of Spohn's proof [45] and it clearly shows the limitation of the method to short macroscopic times.

The source of the $1/n!$ is actually the time-ordered multiple time integration present in (7.7); the total volume of the simplex

$$\{(s_1, s_2, \ldots, s_n) \ : \ 0 \leq s_1 \leq \ldots \leq s_n \leq t\} \tag{7.12}$$

is $t^n/n!$. However, (7.11) contains two such multiple time integrations, but it turns out that only one of them is effective. The source of the competing $n!$ is the pairing that comes from taking the expectation, e.g. in case of Gaussian randomness this appears explicitly in the Wick theorem. The reason for this mechanism is the fact that in quantum mechanics the physical quantities are always quadratic in ψ, thus we have to pair two collision histories in all possible ways even if only one of them has a classical meaning. This additional $n!$ is very typical in all perturbative quantum mechanical argument.

This was only the treatment of the fully expanded terms, the error term in (7.5) needs a separate estimate. As explained in details in Lemma 7.2, here we simply use the unitarity of the full evolution

$$\Big\| \int_0^t e^{i(t-s)H} \underbrace{\int V e^{is_1\Delta} V e^{is_2\Delta} \dots \mathrm{d}\mu_{n,s}(\underline{s})}_{\psi_s^\#} \Big\| \leq \underbrace{t}_{price} \sup_s \|\psi_s^\#\|. \tag{7.13}$$

In this estimate we lose a factor of t, but we end up with a fully expanded term $\psi_s^\#$ which can also be expanded into Feynman graphs.

How to go beyond Spohn's argument that is valid only up to short macroscopic time? Notice that "most" graphs have many "crosses", hence they are expected to be much smaller due to phase decoherence than the naive "one-cross" estimate indicates. One can thus distinguish graphs with one, two and more crossings. From the graphs with more crossings, additional λ^2-factors can be obtained. On the other hand, the combinatorics of the graphs with a few crossings is much smaller than $n!$.

On kinetic scale the following estimate suffices (here N denotes the threshold so that the Duhamel expansion is stopped after N collisions):

$$\mathbf{E}\|\psi_t\|^2 \leq \sum_{n=0}^{N-1} \frac{(\sigma\lambda^2 t)^n}{n!} \Big(\underbrace{1}_{\text{ladder}} + \underbrace{n\lambda^2}_{\text{one cross}} + \underbrace{n!\lambda^4}_{\text{rest}} \Big)$$
$$+ \underbrace{t}_{unit.price} \frac{(\sigma\lambda^2 t)^N}{N!} \Big(\underbrace{1}_{\text{ladder}} + \underbrace{N\lambda^2}_{\text{one cross}} + \underbrace{N^2\lambda^4}_{\text{two cross}} + \underbrace{N!\lambda^6}_{\text{rest}} \Big).$$

Finally, we can optimize

$$N = N(\lambda) \sim (\log\lambda)/(\log\log\lambda)$$

to get the convergence. This estimate gives the kinetic (Boltzmann) limit for all fixed T ($t = T\lambda^{-2}$). This concludes the outline of the proof of Theorem 6.1.

For the proof of Theorem 6.2, i.e. going to diffusive time scales, one needs to classify essentially **all** diagrams; it will not be enough to separate a few crosses and treat all other graphs identically.

8 Key ideas of the proof of the diffusion (Theorems 6.2 and 6.3)

We start with an apologetical remark. For pedagogical reasons, we will present the proof of a mixed model that actually does not exists. The reason is that the dispersion relation is simpler in the continuum case, but the random potential is simpler in the discrete case. The simplicity of the dispersion relation is related to the fact that in continuum the level sets of the dispersion relation, $\{v \,:\, e(v) = e\}$, are convex (spheres), thus several estimates are more effective. The simplicity of the random potential in the discrete case is obvious: no additional form-factor B and no additional Poisson point process needs to be treated. Moreover, the ultraviolet problem

is automatically absent, while in the continuous case the additional large momentum decay needs to be gained from the decay of $\widehat{B}$. Therefore, we will work in the discrete model (i.e. no form factors B, no Poisson points and no large momentum problem), and we will denote the dispersion relation in general by $e(p)$. All formulas hold for both dispersion relations, except some estimates that will be commented on. The necessary modifications from this "fake" model to the two real models are technically sometimes laborous, but not fundamental. We will comment on them at the end of the proof.

8.1 Stopping rules

The Duhamel formula has the advantage that it can be stopped at different number of terms depending on the collision history of each elementary wave function. Thus the threshold N does not have to be chosen uniformly for each term in advance; looking at each term

$$(-i\lambda)^n \int e^{-is_0 H}\, V_{\alpha_1} \cdots V_{\alpha_n} e^{is_n \Delta}\, \psi_0 \, \mathrm{d}\mu_{n,t}(\underline{s})$$

separately, we can decide if we wish to expand $e^{-is_0 H}$ further, or we rather use the unitary estimate (7.13). The decision may depend on the sequence $(\alpha_1, \dots, \alpha_n)$.

Set $K := (\lambda^2 t)\lambda^{-\delta}$ (with some $\delta > 0$) as an absolute upper threshold for the number of collisions in an expanded term (notice that K is much larger typical number of collisions). We stop the expansion if we *either* have reached K expanded potential terms *or* we see a repeated α-label. This procedure has the advantage that repetition terms do not pile up. This stopping rule leads us to the following version of the Duhamel formula:

$$\psi_t = \sum_{n=0}^{K-1} \psi_{n,t}^{nr} + \int_0^t e^{-i(t-s)H} \underbrace{\Big(\psi_{*K,s}^{nr} + \sum_{n=0}^{K} \psi_{*n,s}^{rep} \Big)}_{=:\psi_s^{err}} \mathrm{d}s \tag{8.1}$$

$$\text{with} \qquad \psi_{*n,t}^{nr} := \sum_{A\,:\,\text{nonrep.}} \psi_{*A,t}, \qquad \psi_{*n,s}^{rep} := \sum_{\substack{|A|=n \\ \text{first rep. at } \alpha_n}} \psi_{*A,s} \tag{8.2}$$

Here ψ^{nr} contains collision histories with no repetition, while the repetition terms ψ^{rep} contain exactly one repetition at the last obstacle, since once such a repetition is found, the expansion is stopped. In particular, the second sum above runs over $A = (\alpha_1, \dots \alpha_n)$ where $\alpha_n = \alpha_j$ for some $j < n$ and this is the first repetition. Actually the precise definition of the elementary wave functions in (8.2) is somewhat different from (7.6) (this fact is indicated by the stars), because the expansion starts with a potential term while ψ_A in (7.6) starts with a free evolution.

The backbone of our argument is the following three theorems:

Theorem 8.1. [Error terms are negligible] *We have the following estimate*

$$\sup_{s \le t} \mathbf{E} \|\psi_s^{err}\| = o(t^{-2}).$$

In particular, by the unitarity estimate this trivially implies that

$$\mathbf{E}\Big\| \int_0^t e^{-i(t-s)H} \psi_s^{err} \mathrm{d}s \Big\|^2 = o(1). \tag{8.3}$$

Theorem 8.2. [Only the ladder contributes] *For $n \leq K = (\lambda^2 t)\lambda^{-\delta} = O(\lambda^{-\kappa-\delta})$ we have*

$$\mathbf{E}\|\psi_{n,t}^{nr}\|^2 = Val(id_n) + o(1) \tag{8.4}$$

$$\mathbf{E}\, W_{\psi_{n,t}^{nr}} = Val_{Wig}(id_n) + o(1). \tag{8.5}$$

Here $\mathrm{Val}_{\mathrm{Wig}}$ denotes the value of the Feynman graphs that obtained by expanding the Wigner transform instead of the L^2-norm of ψ_t. Note that (8.5) follows from (8.4) by (the proof of) Lemma 7.1.

Theorem 8.3. [Wigner transform of the main term] *The total contribution of the ladder terms up to K collisions,*

$$\sum_{n=0}^{K} Val_{Wig}(id_n),$$

satisfies the heat equation.

In the following sections we will focus on the non-repetition terms with $n \leq K$, i.e. on Theorem 8.2. We will only discuss the proof for the L^2-norm (8.4), the proof of (8.5) is a trivial modification. Theorem 8.1 is a fairly long case by case study, but it essentially relies on the same ideas that are behind Theorem 8.2. Some of these cases are sketched in Section 10. Finally, Theorem 8.3 is a non-trivial but fairly explicit calculation that we sketch in Section 11.

8.2 Feynman diagrams in the momentum-space formalism

Now we express the value of a Feynman diagram explicitly in momentum representation. We recall the formula (7.6) and Fig. 13 for the representation of the collision history of an elementary wave function. We will express it in momentum space. This means that we assign a running momentum, $p_0, p_1, \ldots p_n$ to each edge of the graph in Fig. 13. Note that we work on the lattice, so all momenta are on the torus, $p \in \mathbf{T}^d = [-\pi, \pi]^d$; in the continuum model, the momenta would run over all $\mathbb{R}^d$. The propagators thus become multiplications with $e^{-is_j e(p_j)}$ and the potentials become convolutions:

$$\begin{aligned}\widehat{\psi}_{A,t}(p) &= \int_{(\mathbf{T}^d)^n} \prod_{j=1}^{n} \mathrm{d}p_j \int \mathrm{d}\mu_{n,t}(\underline{s})\, e^{-is_0 e(p)} \widehat{V}_{\alpha_1}(p-p_1) e^{-is_1 e(p_1)} \widehat{V}_{\alpha_2}(p_1-p_2) \ldots \widehat{\psi}_0(p_n) \\ &= \underbrace{e^{\eta t}}_{\eta:=1/t} \int_{(\mathbf{T}^d)^n} \prod_{j=1}^{n} \mathrm{d}p_j \int_{-\infty}^{\infty} \mathrm{d}\alpha\, e^{-i\alpha t} \prod_{j=1}^{n} \frac{1}{\alpha - e(p_j) + i\eta} \widehat{V}_{\alpha_j}(p_{j-1}-p_j) \widehat{\psi}_0(p_n). \end{aligned} \tag{8.6}$$

In the second step we used the identity

$$\begin{aligned}&\int_{\mathbb{R}_+^{n+1}} \delta\Big(t - \sum_{j=0}^{n} s_j\Big) \mathrm{d}s_0 \mathrm{d}s_1 \ldots \mathrm{d}s_n\, e^{-is_0 e(p_0)} e^{-is_1 e(p_1)} \ldots e^{-is_n e(p_n)} \\ &\qquad = e^{i\eta t} \int_{\mathbb{R}} \mathrm{d}\alpha\, e^{-i\alpha t} \prod_{j=1}^{n} \frac{1}{\alpha - e(p_j) + i\eta} \end{aligned} \tag{8.7}$$

that holds for any $\eta > 0$ (we will choose $\eta = 1/t$ in the applications to neutralize the exponential prefactor). This identity can be obtained from writing

$$\delta\Big(t - \sum_{j=0}^{n} s_j\Big) = \int e^{-i\alpha(t-\sum_{j=0}^{n} s_j)} \mathrm{d}\alpha$$

and integrating out the times. [Note that the letter α is used here and in the sequel as the conjugate variable to time, while previously α_j denoted obstacles. We hope this does not lead to confusion, unfortunately this convention was adopted in our papers as well.]

The L^2-norm is computed by doubling the formula (8.6). We will use the letters $p_1, \dots p_n$ for the momenta in ψ_A and primed momenta for ψ_B:

$$\mathbf{E}\|\psi_t^{nr}\|^2 = \mathbf{E}\Big\| \sum_{A\,:\,\text{nonrep}} \psi_A \Big\|^2 = \sum_{A,B\ \text{nonrep}} \mathbf{E}\langle \psi_A, \psi_B\rangle = \sum_n \sum_{\pi\in\Pi_n} \mathrm{Val}(\pi),$$

$$\mathrm{Val}(\pi) := e^{2\eta t} \int_{-\infty}^{\infty} \mathrm{d}\alpha \mathrm{d}\beta\ e^{i(\alpha-\beta)t} \prod_{j=0}^{n} \frac{1}{\alpha - e(p_j) - i\eta} \frac{1}{\beta - e(p_j') + i\eta} \tag{8.8}$$
$$\times\, \lambda^{2n} \prod_{j=1}^{n} \delta\Big((p_{j-1} - p_j) - (p'_{\pi(j)-1} - p'_{\pi(j)})\Big) |\widehat{\psi}_0(p_n)|^2 \delta(p_n - p_n') \prod_{j=0}^{n} \mathrm{d}p_j \mathrm{d}p_j'.$$

After the pairing, here we computed explicitly the pair expectations:

$$\mathbf{E} \sum_{\alpha\in\mathbb{Z}^d} \overline{\widehat{V}_\alpha(p)} \widehat{V}_\alpha(q) = \sum_{\alpha\in\mathbb{Z}^d} e^{i\alpha(p-q)} = \delta(p-q), \tag{8.9}$$

and we also used that only terms with $B = \pi(A)$ are nonzero. This formula holds for the simplest random potential (5.4) in the discrete model; for the continuum model (5.5) we have

$$\mathbf{E} \sum_{\alpha} \overline{\widehat{V}_\alpha(p)} \widehat{V}_\alpha(q) = |\widehat{B}(p)|^2 \delta(p-q) \tag{8.10}$$

According to the definition of V_ω in the continuum model (5.6), α labels the realizations of the Poisson point process in the last formula (8.10).

The presence of the additional delta function, $\delta(p_n - p_n')$, in (8.8) is due to the fact that we compute the L^2-norm. We remark that the analogous formula in the expansion for Wigner transform differs only in this factor; the arguments of the two wave functions in the momentum representation of the Wigner transform are shifted by a fixed value ξ (4.3), thus the corresponding delta function will be $\delta(p_n - p_n' - \xi)$.

A careful reader may notice that the non-repetition condition on A imposes a restriction on the summation in (8.9); in fact the summation for α_j is not over the whole $\mathbb{Z}^d$, but only for those elements that are distinct from the other α_i's. These terms we will add to complete the sum (8.9) and then we will estimate their effect separately. They correspond to higher order cumulants and their contribution is negligible, but technically they cause serious complications; essentially a complete cumulant expansion needs to be organized. We will shortly comment on them in Section 10.1; we will neglect this issue for the moment.

This calculation gives that

$$\mathbf{E}\|\psi_t^{nr}\|^2 = \sum_n \sum_{\pi\in S_n} \mathrm{Val}(\pi)$$

as it is pictured in Fig. 14. More concisely,

$$\mathrm{Val}(\pi) = \lambda^{2n} e^{2\eta t} \int \mathrm{d}\mathbf{p}\mathrm{d}\mathbf{p}'\mathrm{d}\alpha\mathrm{d}\beta\; e^{it(\alpha-\beta)} \prod_{j=0}^{n} \frac{1}{\alpha - e(p_j) - i\eta}\, \frac{1}{\beta - e(p_j') + i\eta}\, \Delta_\pi(\mathbf{p},\mathbf{p}')\, |\widehat{\psi}_0(p_n)|^2, \tag{8.11}$$

where $\mathbf{p}$ and $\mathbf{p}'$ stand for the collection of integration momenta and

$$\Delta_\pi(\mathbf{p},\mathbf{p}') := \delta(p_n - p_n') \prod_{j=1}^{n} \delta\Big((p_{j-1} - p_j) - (p'_{\pi(j)-1} - p'_{\pi(j)})\Big)$$

contains the product of all delta functions. These delta functions can be obtained from the graph: they express the Kirchoff law at each pair of vertices, i.e. the signed sum of the four momenta attached to any paired vertices must be zero. It is one of the main advantages of the Feynman graph representation that the complicated structure of momentum delta functions can be easily read off from the graph.

We remark that in (8.11) we omitted the integration domains; the momentum integrations runs through the momentum space, i.e. each p_j and p_j' is integrated over $\mathbb{R}^d$ or $\mathbf{T}^d$, depending whether we consider the continuum or the lattice model. The $\mathrm{d}\alpha$ and $\mathrm{d}\beta$ integrations always run through the reals. We will adopt this short notation in the future as well:

$$\int \mathrm{d}\mathbf{p}\mathrm{d}\mathbf{p}'\mathrm{d}\alpha\mathrm{d}\beta = \int_{(\mathbf{T}^d)^{n+1}} \prod_{j=0}^{n} \mathrm{d}p_j \int_{(\mathbf{T}^d)^{n+1}} \prod_{j=0}^{n} \mathrm{d}p_j' \int_{\mathbb{R}} \mathrm{d}\alpha \int_{\mathbb{R}} \mathrm{d}\beta.$$

We will also often use momentum integrations without indicating their domains, which is always $\mathbb{R}^d$ or $\mathbf{T}^d$.

8.3 Lower order examples

In this section we compute explicitly a few low order diagrams for the Wigner transform. Recalling that $\Phi_N = \sum_{n=0}^{N-1} \psi^{(n)}$ represents the fully expanded terms and recalling the Wigner transform in momentum space (4.3), we can write the (rescaled) Wigner transform of Φ_N as follows:

$$\widehat{W}^{\varepsilon}_{\Phi_N}(\xi, v) = \sum_{n'=0}^{N-1} \sum_{n=0}^{N-1} \overline{\widehat{\psi}_t^{(n')}\big(v - \frac{\varepsilon}{2}\xi\big)}\, \widehat{\psi}_t^{(n)}\big(v + \frac{\varepsilon}{2}\xi\big) =: \sum_{n,n'=0}^{N-1} \widehat{W}^{\varepsilon}_{n,n',t}(\xi, v).$$

Set $k_n = v + \frac{\varepsilon}{2}\xi$ and $k'_{n'} := v - \frac{\varepsilon}{2}\xi$, then we can write

$$\widehat{\psi}_t(k_n) = (-i\lambda)^n \int \mathrm{d}\mu_{n,t}(\underline{s}) e^{-is_n e(k_n)} \int \prod_{j=0}^{n-1} \Big(\mathrm{d}k_j\; e^{-is_j e(k_j)}\Big) \prod_{j=1}^{n} \widehat{V}(k_j - k_{j-1}) \widehat{\psi}_0(k),$$

which can be doubled to obtain $\widehat{W}^{\varepsilon}_{n,n',t}$. Notice that we shifted all integration variables, i.e. we use $k_j = v_j + \frac{\varepsilon}{2}\xi$.

Case 1: $n = n' = 0$. In this case there is no need for taking expectation and we have

$$\widehat{W}^{\varepsilon}_{n,n',t}(\xi, v) = e^{it(e(k_0')-e(k_0))}\widehat{\psi}_0(k_0)\overline{\widehat{\psi}_0(k_0')} = e^{it(e(k_0')-e(k_0))}\widehat{W}^{\varepsilon}_0(\xi, v)$$

where W^{ε}_0 is the Wigner transform of the initial wave function ψ_0. Suppose, for simplicity, that the initial wave function is unscaled, i.e. it is supported near the origin in position space, uniformly in λ. Then it is easy to show that its rescaled Wigner transform converges to the Dirac delta measure in position space:

$$\lim_{\varepsilon\to 0} W^{\varepsilon}_{\psi_0}(X, v)\mathrm{d}X\mathrm{d}v = \delta(X)|\widehat{\psi}_0(v)|^2\mathrm{d}X\mathrm{d}v$$

since for any fixed ξ,

$$\widehat{\psi}_0(k_0)\overline{\widehat{\psi}_0(k_0')} \to |\widehat{\psi}_0(v)|^2 \tag{8.12}$$

as $\varepsilon \to 0$. In the sequel we will use the continuous dispersion relation $e(k) = \frac{1}{2}k^2$ for simplicity because it produces explicit formulas. In the general case, similar formulas are obtained by Taylor expanding $e(k)$ in ε up to the first order term and higher order terms are negligible but need to be estimated.

Since

$$e(k_0') - e(k_0) = \frac{1}{2}\Big(v - \frac{\varepsilon}{2}\xi\Big)^2 - \frac{1}{2}\Big(v + \frac{\varepsilon}{2}\xi\Big)^2 = -\varepsilon v\cdot\xi,$$

we get that in the kinetic limit, when $\lambda^2 t = T$ is fixed and $\varepsilon = \lambda^2$

$$\lim_{\varepsilon\to 0}\int \widehat{J}(\xi, v)\widehat{W}^{\varepsilon}_{n,n',t}(\xi, v)\mathrm{d}\xi\mathrm{d}v = \int \widehat{J}(\xi, v)|\widehat{\psi}_0(v)|^2 e^{-iTv\cdot\xi}\mathrm{d}\xi\mathrm{d}v = \int J(Tv, v)|\widehat{\psi}_0(v)|^2\mathrm{d}v,$$

by using dominated convergence and (8.12). The last formula is the weak formulation of the evolution of initial phase space measure $\delta(X)|\widehat{\psi}_0(v)|^2\mathrm{d}X\mathrm{d}v$ under the free motion along straight lines.

Case 2: $n = 0$, $n' = 1$ or $n = 1$, $n' = 0$. Since there is one single potential in the expansion, these terms are zero after taking the expectation:

$$\mathbf{E}\,\widehat{W}^{\varepsilon}_{1,0,t}(\xi, v) = \mathbf{E}\,\widehat{W}^{\varepsilon}_{0,1,t}(\xi, v) = 0.$$

Case 3: $n = n' = 1$. We have

$$\begin{aligned}\widehat{W}^{\varepsilon}_{1,1,t}(\xi, v) =& (-i\lambda)\int_0^t \mathrm{d}s\; e^{-i(t-s)e(k_1)}\int \mathrm{d}k_0\widehat{V}(k_1 - k_0)e^{-ise(k_0)}\widehat{\psi}_0(k_0)\\ &\times (i\lambda)\int_0^t \mathrm{d}s'\; e^{i(t-s')e(k_1')}\int \mathrm{d}k_0'\overline{\widehat{V}(k_1' - k_0')}e^{is'e(k_0')}\overline{\widehat{\psi}_0(k_0')}.\end{aligned} \tag{8.13}$$

The expectation value acts only on the potentials, and we have (using now the correct continuum potential)

$$\mathbf{E}\widehat{V}(k_1 - k_0)\overline{\widehat{V}(k_1' - k_0')} = |\widehat{B}(k_1 - k_0)|^2\delta\big(k_1' - k_0' - (k_1 - k_0)\big).$$

After collecting the exponents, we obtain

$$\mathbf{E}\widehat{W}^{\varepsilon}_{1,1,t}(\xi,v) = \lambda^2 \int \mathrm{d}k_0 \mathrm{d}k_0' |\widehat{B}(k_1-k_0)|^2 \delta\big(k_1'-k_0'-(k_1-k_0)\big) \\ \times e^{it\left(e(k_1')-e(k_1)\right)} Q_t\big(e(k_1)-e(k_0)\big)\overline{Q}_t\big(e(k_1')-e(k_0')\big)\widehat{\psi}_0(k_0)\overline{\widehat{\psi}_0(k_0')},$$

where we introduced the function

$$Q_t(a) := \frac{e^{ita}-1}{ia}.$$

Consider first the case $\xi = 0$, i.e. when the Wigner transform becomes just the momentum space density. Then $v = k_1 = k_1'$ and thus $k_0 = k_0'$ by the delta function, so $e(k_1) - e(k_0) = e(k_1') - e(k_0')$ and we get

$$\mathbf{E}\widehat{W}^{\varepsilon}_{1,1,t}(0,v) = \lambda^2 \int \mathrm{d}k_0 |\widehat{B}(k_1-k_0)|^2 |\widehat{\psi}_0(k_0)|^2 \big|Q_t\big(e(k_1)-e(k_0)\big)\big|^2.$$

Clearly

$$t^{-1}|Q_t(A)|^2 \to 2\pi\delta(A)$$

as $t \to \infty$ (since $\int \left(\frac{\sin x}{x}\right)^2 \mathrm{d}x = \pi$), so we obtain that in the limit $t \to \infty$, $\lambda^2 t = T$ fixed,

$$\mathbf{E}\widehat{W}^{\varepsilon}_{1,1,t}(0,v) \to T \int \mathrm{d}k_0 |\widehat{B}(k_1-k_0)|^2 2\pi\delta\big(e(k_1)-e(k_0)\big)|\widehat{\psi}_0(k_0)|^2 = T\int \mathrm{d}k_0 \sigma(k_1,k_0)|\widehat{\psi}_0(k_0)|^2 \tag{8.14}$$

(recall $k_1 = v$ in this case), where the collision kernel is given by

$$\sigma(k_1,k_0) := |\widehat{B}(k_1-k_0)|^2 2\pi\delta\big(e(k_1)-e(k_0)\big).$$

A similar but somewhat more involved calculation gives the result for a general ξ, after testing it against a smooth function J. Then it is easier to use (8.13) directly, and we get, after taking expectation and changing variables,

$$\mathbf{E}\widehat{W}^{\varepsilon}_{1,1,t}(\xi,v) = \lambda^2 \int_0^t \mathrm{d}s \int_0^t \mathrm{d}s' \int \mathrm{d}v_0\, e^{i\Phi/2} |\widehat{B}(v-v_0)|^2 \widehat{W}^{\varepsilon}_0(\xi,v_0)$$

with a total phase factor

$$\Phi := -(t-s)\Big(v+\frac{\varepsilon}{2}\xi\Big)^2 - s\Big(v_0+\frac{\varepsilon}{2}\xi\Big)^2 + (t-s')\Big(v-\frac{\varepsilon}{2}\xi\Big)^2 + s'\Big(v_0-\frac{\varepsilon}{2}\xi\Big)^2$$

$$= (s-s')(v^2-v_0^2) - 2\varepsilon\Big[\big(t-\frac{s+s'}{2}\big)v - \frac{s+s'}{2}v_0\Big]\cdot\xi.$$

After changing variables: $b = s - s'$ and $T_0 = \varepsilon\frac{s+s'}{2}$, we notice that the first summand gives

$$\int \mathrm{d}b\, e^{ib(v^2-v_0^2)} = 2\pi\delta(v^2-v_0^2), \tag{8.15}$$

and from the second one we have

$$\mathbf{E}\widehat{W}^{\varepsilon}_{1,1,t}(\xi,v)=\int_0^T \mathrm{d}T_0\int \mathrm{d}v_0\; e^{-i\left[(T-T_0)v-T_0v_0\right]\cdot\xi}\sigma(v,v_0)\widehat{W}^{\varepsilon}_0(\xi,v_0).$$

These steps, especially (8.15), can be made rigorous only if tested against a smooth function J, so we have

$$\mathbf{E}\langle J,W_{1,1,t}\rangle\to\int_0^T \mathrm{d}T_0\int \mathrm{d}v_0 J\Big((T-T_0)v-T_0v_0,v\Big)\sigma(v,v_0)|\widehat{\psi}_0(v_0)|^2.$$

The dynamics described in the first variable of J is a free motion with velocity v_0 up to time T_0, then a fictitious collision happens that changes the velocity from v_0 to v (and this process is given by the rate $\sigma(v,v_0)$) and then free evolution continues during the remaining time $T-T_0$ with velocity v. This is exactly the one "gain" collision term in the Boltzmann equation.

Case 4: $n=0$, $n'=2$ or $n=2$, $n'=0$.
The calculation is similar to the previous case, we just record the result:

$$\mathbf{E}\widehat{W}_{2,0,t}(\xi,v)=-\lambda^2t\widehat{\psi}_0(k_1)\overline{\widehat{\psi}_0(k_1')}\;e^{-it\left(e(k_1)-e(k_1')\right)}\int \mathrm{d}q|\widehat{B}(k_1-q)|^2tR\big(t(e(k_1)-e(q))\big),$$

where

$$R(u)=\frac{e^{iu}-iu-1}{u^2}.$$

Simple calculation shows that $tR(tu)\to\pi\delta(u)$ as $t\to\infty$. Thus we have

$$\mathbf{E}\widehat{W}_{2,0,t}(\xi,v)\to-\frac{T}{2}|\widehat{\psi}_0(v)|^2e^{-iTv\cdot\xi}\sigma(v),$$

where we defined

$$\sigma(v):=\int\sigma(v,q)\mathrm{d}q.$$

Similar result holds for $\widehat{W}_{0,2,t}$ and combining it with (8.14), we obtain the conservation of the L^2-norm up to second order in λ, since with $\xi=0$ we have

$$\int \mathrm{d}v\Big[\mathbf{E}\widehat{W}_{1,1,t}(0,v)+\mathbf{E}\widehat{W}_{2,0,t}(0,v)+\mathbf{E}\widehat{W}_{0,2,t}(0,v)\Big]=0.$$

For general ξ and after testing against a smooth function, we obtain

$$\mathbf{E}\langle J,W_{2,0,t}\rangle\to-\frac{T}{2}\int \mathrm{d}vJ(Tv,v)\sigma(v)|\widehat{\psi}_0(v)|^2$$

which corresponds to the loss term in the Boltzmann equation.

Apart from revealing how the Boltzmann equation emerges from the quantum expansion, the above calculation carries another important observation. Notice that terms with $n\neq n'$ did give rise to non-negligible contributions despite the earlier statement (7.8) that $n=n'$ is forced by the non-repetition rule (which is correct) and that repetitive collision sequences are negligible (which is, apparently, not correct). The next section will explain this.

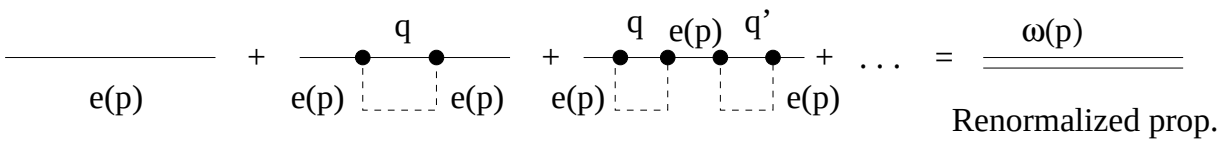

Figure 17: Schematic picture of the renormalization

8.4 Self-energy renormalization

The previous simple explicit calculations showed that not all repetition terms are neglible, in fact immediate recollisions are of order one and they – but only they – have to be treated differently. We say that a collision sequence $(\alpha_1, \alpha_2, \ldots)$ has an *immediate recollision* if $\alpha_i = \alpha_{i+1}$ for some i. This modification must also be implemented in the stopping rule, such that immediate recollisions do not qualify for recollisions for the purpose of collecting reasons to stop the expansion (see Section 8.1).

Fortunately immediate repetitions appear very locally in the Feynman graph and they can be resummed according to the schematic picture on Fig. 17. The net effect is that the free propagator $e(p) = \frac{1}{2}\,p^2$ needs to be changed to another propagator $\omega(p)$ that differs from $e(p)$ by an $O(\lambda^2)$ amount. More precisely, using the time independent formalism, second line of (8.6), we notice that the first term in Fig. 17 is represented by

$$\frac{1}{\alpha - e(p) - i\eta} \tag{8.16}$$

(which is also often called propagator). The second term carries a loop integration (momentum q), and its contribution is

$$\frac{1}{(\alpha - e(p) - i\eta)^2} \int \frac{\lambda^2 |\widehat{B}(p-q)|^2}{\alpha - e(q) - i\eta}\,\mathrm{d}q.$$

This formula is written for the continuum model and the integration domain is the whole momentum space $\mathbb{R}^d$. For the lattice model $\widehat{B}$ is absent and the integration is over $\mathbf{T}^d$.

The third term carries two independent loop integration (momenta q and q'), and its contribution is

$$\frac{1}{(\alpha - e(p) - i\eta)^3} \int \frac{\lambda^2 |\widehat{B}(p-q)|^2 \mathrm{d}q}{\alpha - e(q) - i\eta} \int \frac{\lambda^2 |\widehat{B}(p-q')|^2 \mathrm{d}q'}{\alpha - e(q') - i\eta}.$$

Setting

$$\Theta_\eta(p, \alpha) := \frac{|\widehat{B}(p-q)|^2 \mathrm{d}q}{\alpha - e(q) - i\eta}$$

and noticing that due to the almost singularities of the $(\alpha - e(p) - i\eta)^{-k}$ prefactors the main contribution comes from $\alpha \sim e(p)$, we can set

$$\theta(p) := \lim_{\eta \to 0+} \Theta\big(p, e(p)\big). \tag{8.17}$$

(Similar, in fact easier formulas hold for the lattice model). Therefore, modulo negligible errors, the sum of the graphs in Fig. 17 give rise to the following geometric series:

$$\frac{1}{\alpha - e(p) - i\eta} + \frac{\lambda^2\theta(p)}{(\alpha - e(p) - i\eta)^2} + \frac{\big(\lambda^2\theta(p)\big)^2}{(\alpha - e(p) - i\eta)^3} + \ldots = \frac{1}{\alpha - \big(e(p) + \lambda^2\theta(p)\big) - i\eta}$$

This justifies to define the *renormalized propagator*

$$\omega(p) := e(p) + \lambda^2\theta(p)$$

and the above calculation indicates that all immediate recollisions can be taken into account by simply replacing $e(p)$ with $\omega(p)$. This is in fact can be rigorously proved up to the leading order we are interested.

An alternative way to see the renormalization is to reorganize how the original Hamiltonian is split into main and perturbation terms:

$$H = \underbrace{e(p) + \lambda^2\theta(p)}_{\omega(p)} + \lambda V - \lambda^2\theta(p).$$

The precise definition of the correction term $\theta(p)$ is determined by the following self-consistent equation:

$$\theta(p) := \int \frac{\mathrm{d}q}{\omega(p) - \omega(q) + i0},$$

the formula (8.17) is in fact only the solution to this equation up to order λ^2, but for our purposes such precision is sufficient. The imaginary part of θ can also be computed as

$$\sigma(p) := \eta \int \frac{\mathrm{d}q}{|\omega(p) - \omega(q) + i\eta|^2} \to \mathrm{Im}\theta(p) \qquad \eta \to 0$$

and notice that it is not zero. In particular, the renormalization *regularizes the propagator*: while the trivial supremum bound on the original propagator is

$$\sup_p \Big|\frac{1}{\alpha - e(p) - i\eta}\Big| \le \eta^{-1}$$

the similar bound on the renormalized propagator is much better:

$$\Big|\frac{1}{\alpha - \omega(p) - i\eta}\Big| \le \frac{1}{\lambda^2 + \eta}. \tag{8.18}$$

Strictly speaking, this bound does not hold if $p \approx 0$ since $\mathrm{Im}\,\theta(p)$ vanishes at the origin, but such regime in the momentum space has a small volume, since it is given by an approximate point singularity, while the (almost) singularity manifold of (8.16) is large, it has codimension one.

The bound (8.18) will play a crucial role in our estimates. Recall that due to the exponential prefactor $e^{2\eta t}$ in (8.11), eventually we will have to choose $\eta \sim 1/t$. Thus in the diffusive scaling, when $t \gg \lambda^{-2}$, the bound (8.18) is a substantial improvement.

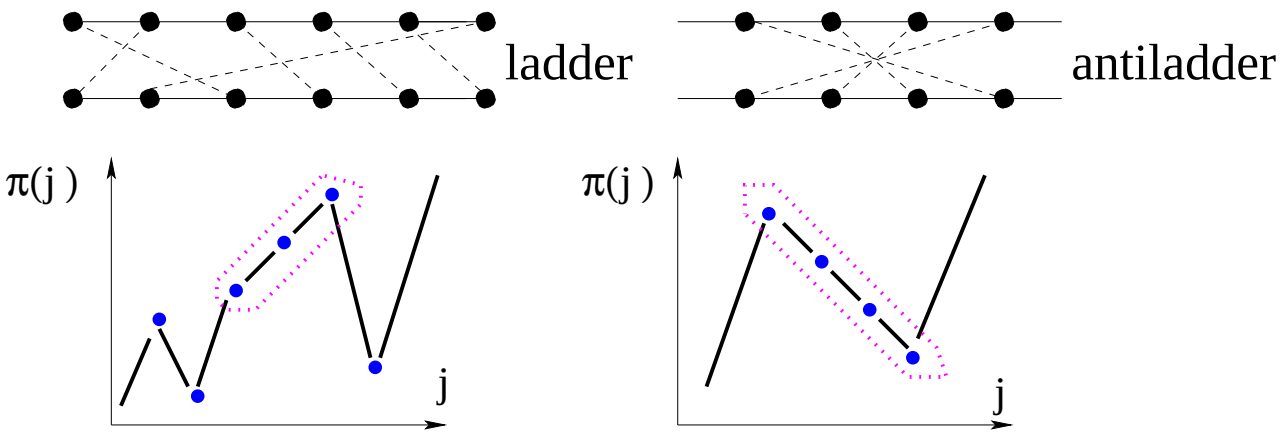

Figure 18: Ladder and antiladder

The precise formulas are not particularly interesting; the main point is that after renormalization: only the ladder has classical contribution and gives the limiting equation. If one uses the renormalized propagators, then one can assume that no immediate repetitions occur in the expansion (in practice they do occur, but they are algebraically cancelled out by the renormalization of the propagator). Only after this renormalization will the value of the ladder graph given in Fig. 15 be correct with the exponential damping factor. From now on we will assume that the renormalization is performed, the propagator is $\omega(p)$ and there is no immediate recollision. In particular, the statements of the key Theorems 8.1 and 8.2 are understood *after this renormalization.*

8.5 Control of the crossing terms

The key to the proof of Theorem 8.2 is a good classification of all Feynman diagrams based upon the complexity of the permutation $\pi : A \to B$. This complexity is expressed by a degree $d(\pi)$ that is defined, temporarily, as follows. We point out that this is a little simplified definition, the final definition is a bit more involved and is given later in Definition 9.1.

Definition 8.4. *Given a permutation $\pi \in S_n$ on $\{1, 2, \ldots, n\}$, an index i is called* **ladder index** *if $|\pi(i) - \pi(i-1)| = 1$ or $|\pi(i) - \pi(i+1)| = 1$. The* **degree of the permutation** *π is defined as*

$$d(\pi) = \#\{\textit{non-ladder indices}\}. \tag{8.19}$$

Two examples are shown on Fig. 18. The top pictures show the pairing in the Feynman diagrams, the bottom pictures show the graph of the permutation as a function

$$\pi : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\},$$

where the set $\{1, 2, \ldots, n\}$ is naturally embedded into the reals and we conveniently joined the discrete points of the graph of π. The dotted region encircles the ladder indices: notice that a long antiladder also has many ladder indices. This indeed shows that it is not really the number of total crosses in the Feynman graph that is responsible for the smallness of the value

of the Feynman graph, rather the disordered structure and that is more precisely expressed by the non-ladder indices.

The main philosophy is that most permutations have a high degree (combinatorial complexity). The key estimates are the following two lemmas. The first lemma estimates the number of permutations with a given degree; the proof is an elementary combinatorial counting and will not be presented here.

Lemma 8.5. *The number of permutations with a given degree is estimated as*

$$\#\{\pi \ : \ d(\pi) = d\} \leq (Cn)^d.$$

The second lemma is indeed the hard part of the proof; it claims that the value of the Feynman graph decreases polynomially (in λ) as the degree increases. Thus we can *gain a* λ *factor per each non-ladder vertex* of the Feynman graph.

Lemma 8.6. *There exists some positive* κ*, depending on the dimension* d*, such that*

$$Val(\pi) \leq (C\lambda)^{\kappa d(\pi)}. \tag{8.20}$$

Combining these two Lemmas, we can easily control the series $\sum_\pi \mathrm{Val}(\pi)$:

$$\sum_{\pi \in S_n} \mathrm{Val}(\pi) = \sum_{d=0}^{\infty} \sum_{\pi : d(\pi)=d} \mathrm{Val}(\pi) = \sum_d C^d n^d \lambda^{\kappa d} < \infty$$

if $n \leq K \sim \lambda^{-\kappa}$ (assume $\delta = 0$ for simplicity). Since $n \sim \lambda^2 t$, get convergence for $t \leq c\lambda^{-2-\kappa}$, i.e. the κ from Lemma 8.6 determines the time scale for which our proof is valid.

We remark that, although for technical reasons we can prove (8.20) only for very small κ, it should be valid up to $\kappa = 2$ but not beyond. To see this, recall that the best possible estimate for a single Feynman graph is $O(\lambda^{2n})$ and their total number is $n!$. Since

$$\lambda^{2n} n! \approx (\lambda^2 n)^n \approx (\lambda^4 t)^n,$$

the summation over all Feynman graphs will diverge if $t \geq \lambda^{-4}$. This means that this method is limited up to times $t \ll \lambda^{-4}$. Going beyond $t \sim \lambda^{-4}$ requires a second resummation procedure, namely the resummation of the so-called *four-legged* subdiagrams. In other words, one cannot afford to estimate each Feynman graph individually, a certain cancellation mechanism among them has to be found. Similar resummations have been done in many-body problems in euclidean (imaginary time) theories but not in real time. In the current problem it is not clear even on the intuitive level which diagrams cancel each other.

8.6 An example

In this very concrete example we indicate why the cross is smaller than the ladder by a factor λ^2, i.e. we justify the first estimate in Fig. 16. We recall that the value of a Feynman graph of order n is

$$\mathrm{Val}(\pi) = \lambda^{2n} e^{2\eta t} \int \mathrm{d}\mathbf{p}\mathrm{d}\mathbf{p}'\mathrm{d}\alpha\mathrm{d}\beta \; e^{it(\alpha-\beta)} \prod_{j=0}^{n} \frac{1}{\alpha - \overline{\omega}(p_j) - i\eta} \; \frac{1}{\beta - \omega(p'_j) + i\eta} \; \Delta_\pi(\mathbf{p}, \mathbf{p}') \ , \tag{8.21}$$

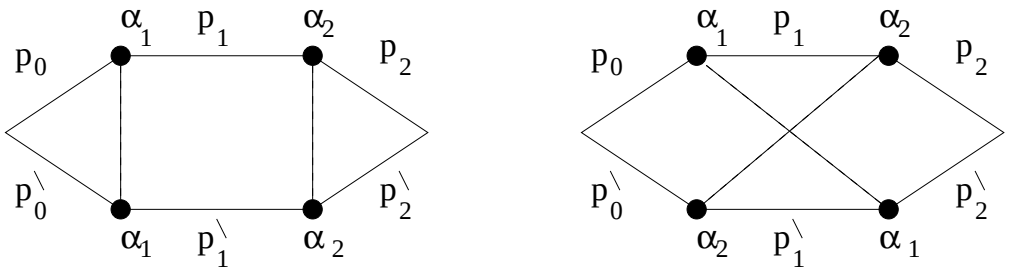

Figure 19: Ladder and cross

where the delta function comes from momentum conservation. Notice that the integrand is a function that are almost singular on the level sets of the dispersion relation, i.e. on the manifolds

$$\alpha = \mathrm{Re}\,\omega(p_j) = e(p_j) + \lambda^2 \mathrm{Re}\,\theta(p_j), \qquad \beta = \mathrm{Re}\,\omega(p'_j) = e(p'_j) + \lambda^2 \mathrm{Re}\,\theta(p'_j).$$

For the typical values of α and β, these are manifolds of codimension d in the $2(n+1)d$ dimensional space of momenta. The singularities are regularized by the imaginary parts of the denominators, $\mathrm{Im}\,\omega(p_j) + \eta$, which are typically positive quantities of order $O(\lambda^2 + \eta)$. The main contribution to the integral comes from regimes of integration near these "almost singularity" manifolds. We remark that in the continuum model the momentum space extends to infinity, so in principle one has to control the integrals in the large momentum (ultraviolet) regime as well, but this is only a technicality.

The main complication in evaluating and estimating the integral (8.21) comes from the delta functions in $\Delta_\pi(\mathbf{p}, \mathbf{p}')$ since they may enhance singularity overlaps and may increase the value of the integral. Without these delta functions, each momentum integration $\mathrm{d}p_j$ and $\mathrm{d}p'_j$ could be performed independently and the total value of the Feynman graph would be very small, of order $\lambda^{2n} |\log \lambda|^{2(n+1)}$. It is exactly the overlap of these (almost) singularities, forced by the delta functions, that is responsible for the correct size of the integral.

To illustrate this effect, we consider the simplest example for a cross and compare it with the direct pairing (ladder). The notations are found on Fig. 19. For simplicity, we assume that $\lambda^2 = \eta$, i.e. we are in the kinetic regime and the extra regularization coming from the renormalization, $\mathrm{Im}\,\theta$, is not important.

In case of the ladder, the delta functions force all paired momenta to be exactly the same (assuming that $p_0 = p'_0$ since we compute the L^2-norm of ψ_t), i.e.

$$\text{Ladder} \quad \Longrightarrow \quad p_j = p'_j, \quad \forall j = 0, 1, 2.$$

In contrast, the crossing delta functions yield the following relations:

$$p'_0 = p_0, \quad p'_1 = p_0 - p_1 + p_2, \quad p'_2 = p_2.$$

Now we can compute explicitly:

$$\mathrm{Val}(\text{ladder}) = \lambda^4 e^{2\eta t} \int \mathrm{d}\alpha \mathrm{d}\beta\, e^{i(\alpha-\beta)t} \int \frac{1}{\alpha - \omega(p_0) - i\eta} \frac{1}{\alpha - \omega(p_1) - i\eta} \frac{1}{\alpha - \omega(p_2) - i\eta} \times \frac{1}{\beta - \omega(p_0) + i\eta} \frac{1}{\beta - \omega(p_1) + i\eta} \frac{1}{\beta - \omega(p_2) + i\eta} \mathrm{d}p_0 \mathrm{d}p_1 \mathrm{d}p_2. \quad (8.22)$$

Choosing $\eta = 1/t$, with a simple analysis one can verify that the main contribution comes from the regime where $|\alpha - \beta| \lesssim t^{-1}$, thus effectively all singularities overlap (with a precision η). We have

$$\int \frac{\mathrm{d}p}{|\alpha - \omega(p) + i\eta|^2} \sim \frac{1}{|\mathrm{Im}\,\omega|} \sim \lambda^{-2}. \tag{8.23}$$

The intermediate relations are not completely correct as they stand because ω depends on the momentum and its imaginary part actually vanishes at $p = 0$. However, the volume of this region is small. Moreover, in the continuum case the ultraviolet regime needs attention as well, but in that case the decaying form factor $|\widehat{B}|^2$ is also present. A more careful calculation shows that the final relation in (8.23) is nevertheless correct.

Thus, effectively, we have

$$\mathrm{Val}(\mathrm{Ladder}) \sim \lambda^4(\lambda^{-2})^2 \sim O(1) \tag{8.24}$$

modulo logarithmic corrections. Here (8.23) has been used twice and the last pair of denominators integrated out by $\mathrm{d}\alpha\mathrm{d}\beta$ collecting a logarithmic term:

$$\int \frac{\mathrm{d}\alpha}{|\alpha - \omega(p_0) - i\eta|} = O(|\log \eta|) \tag{8.25}$$

(recall that $\eta = 1/t \sim \lambda^2$). This estimate again is not completely correct as it stands, because the integral in (8.25) is logarithmically divergent at infinity, but the ultraviolet regime is always trivial in this problem. Technically, for example, here one can save a little more α decay from the other denominators in (8.22). Both inequalities (8.23) and (8.25) hold both in the discrete and continuum case. In Appendix A we listed them and some more complicated related estimates more precisely that will also be used.

As a rule of thumb, we should keep in mind that the main contribution in all our integrals come from the regimes where:

(i) The two dual variables to the time are close with a precision $1/t$, i.e.

$$|\alpha - \beta| \leq 1/t;$$

(ii) The momenta are of order one and away from zero (i.e. there is no ultraviolet or infrared issue in this problem);

(iii) The variables α, β are also of order one, i.e. there is no divergence at infinity for their integrations.

By an *alternative argument* one can simply estimate all (but one) β-denominators in (8.22) by an L^∞-bound (8.18), i.e. by $O(\lambda^{-2})$, then integrate out $\mathrm{d}\beta$ and then all α-denominators, using (8.25) and the similar L^1-bound

$$\int \frac{\mathrm{d}p}{|\alpha - \omega(p) - i\eta|} = O(|\log \eta|) \tag{8.26}$$

for the momentum integrals. The result is

$$\mathrm{Val}(\mathrm{Ladder}) \lesssim \lambda^4(\lambda^{-2})^2(\log \lambda)^4 = (\log \lambda)^4$$

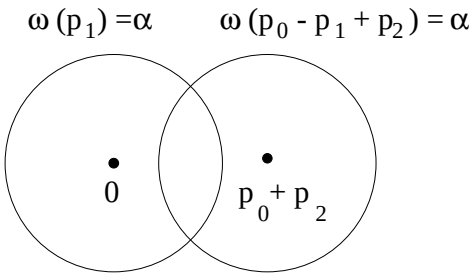

Figure 20: Overlap of the shifted level sets in the p_1-space

which coincides with the previous calculation (8.24) modulo logarithmic terms. In fact, one can verify that these are not only upper estimates for the ladder, but in fact the value of the n-th order ladder is

$$\mathrm{Val}(\mathrm{Ladder}) \sim O_n(1) \tag{8.27}$$

as $\lambda \to 0$, modulo $|\log \lambda|$ corrections.

The similar calculation for the crossed diagram yields:

$$\mathrm{Val}(\mathrm{cross}) = \lambda^4 \int \mathrm{d}\alpha \mathrm{d}\beta \, e^{i(\alpha-\beta)t} \int \frac{1}{\alpha - \omega(p_0) - i\eta} \frac{1}{\alpha - \omega(p_1) - i\eta} \frac{1}{\alpha - \omega(p_2) - i\eta} \times \frac{1}{\beta - \omega(p_0) + i\eta} \frac{1}{\beta - \omega(p_0 - p_1 + p_2) + i\eta} \frac{1}{\beta - \omega(p_2) + i\eta} \mathrm{d}p_0 \mathrm{d}p_1 \mathrm{d}p_2. \tag{8.28}$$

Assuming again that the main contribution is from the regime where $\alpha \sim \beta$ (with precision $1/t$), we notice that spherical singularities of the two middle denominators overlap only at a point singularity

$$\int \mathrm{d}p_1 \frac{1}{|\alpha - \omega(p_1) - i\eta|} \frac{1}{|\alpha - \omega(p_0 - p_1 + p_2) + i\eta|} \sim \frac{1}{|p_0 + p_2| + \lambda^2} \tag{8.29}$$

(see also Fig. 20 and Lemma A.2 for a more precise estimate). In three dimensions the point singularity is harmless (will disappear by the next integration using (A.5) from Lemma A.2), and we thus obtain

$$\mathrm{Val}(\mathrm{cross}) \le \lambda^2 \mathrm{Val}(\mathrm{ladder})$$

(modulo logarithms).

We emphasize that inequality (8.29) in this form holds only for the continuous dispersion relation (see (A.4) in Appendix A), i.e. *if the level sets of the dispersion relation are convex*, since it is relied on the fact that a convex set and its shifted copy overlap transversally, thus a small neighborhood of these sets (where $|\alpha - e(p)|$ is small) have a small intersection. This is wrong for the level sets of the discrete dispersion relation which, for a certain range of α, is not a convex set (see Fig. 21). However, (8.29) holds with an additional factor $\eta^{-3/4}$ on the left hand side (see (A.7)), which is a weaker estimate than in the continuous case but it is still useful because it is stronger than the trivial estimate η^{-1} obtained by by taking the L^∞-norm of one propagator on the left hand side of (8.29).

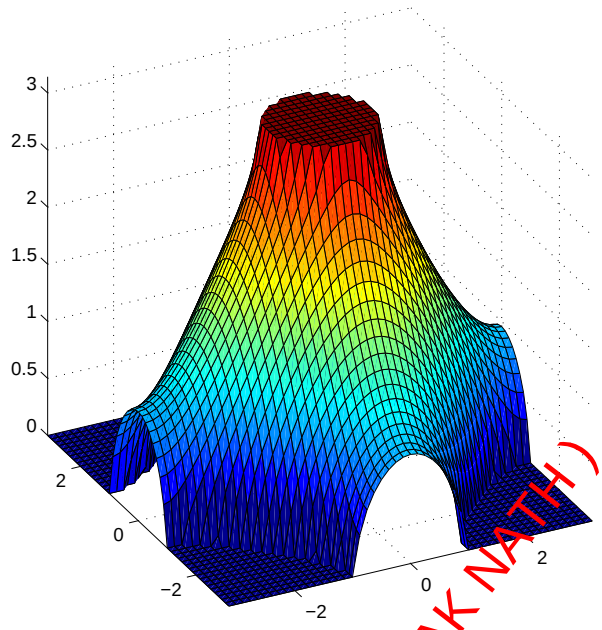

Figure 21: A level set of the discrete dispersion relation $e(p)$ from (5.3)

9 Integration of general Feynman graphs

9.1 Formulas and the definition of the degree

We introduce a different encoding for the permutation π by using permutation matrices. To follow the standard convention of labelling the rows and columns of a matrix by positive natural numbers, we will shift all indices in (8.11) by one, i.e. from now on we will work with the formula

$$\mathrm{Val}(\pi) = \lambda^{2n} e^{2\eta t} \int \mathrm{d}\mathbf{p}\mathrm{d}\mathbf{p}'\mathrm{d}\alpha\mathrm{d}\beta \; e^{it(\alpha-\beta)} \prod_{j=1}^{n+1} \frac{1}{\alpha - \overline{\omega}(p_j) - i\eta} \; \frac{1}{\beta - \omega(p'_j) + i\eta} \; \Delta_\pi(\mathbf{p},\mathbf{p}') \; , \quad (9.1)$$

where

$$\Delta_\pi(\mathbf{p},\mathbf{p}') := \delta(p_{n+1} - p'_{n+1}) \prod_{j=1}^{n} \delta\Big((p_{j+1} - p_j) - (p'_{\pi(j)+1} - p'_{\pi(j)})\Big).$$

We introduce a convenient notation. For any $(n+1)\times(n+1)$ matrix M and for any vector of momenta $\mathbf{p} = (p_1, \dots p_{n+1})$, we let $M\mathbf{p}$ denote the following $(n+1)$-vector of momenta

$$M\mathbf{p} := \Big(\sum_{j=1}^{n+1} M_{1j} p_j, \; \sum_{j=1}^{n+1} M_{2j} p_j, \dots \Big) \; . \quad (9.2)$$

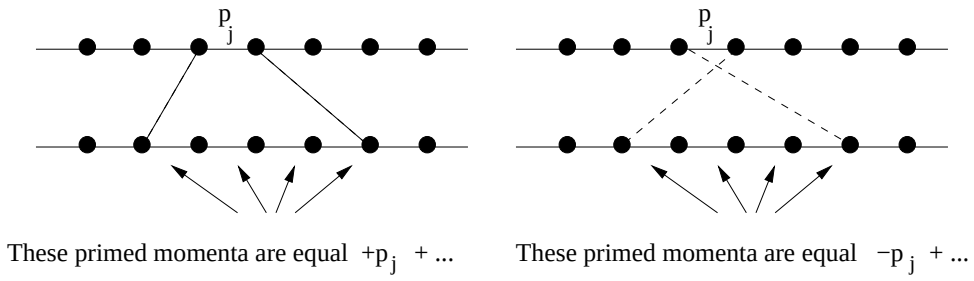

Figure 22: Domain of dependencies of the momenta

The permutation $\pi \in \mathcal{S}_n$ acting on the indices $\{1, 2, \ldots, n\}$ in (9.1) will be encoded by an $(n+1) \times (n+1)$ matrix $M(\pi)$ defined as follows

$$M_{ij}(\pi) := \begin{cases} 1 & \text{if} \quad \widetilde{\pi}(j-1) < i \leq \widetilde{\pi}(j) \\ -1 & \text{if} \quad \widetilde{\pi}(j) < i \leq \widetilde{\pi}(j-1) \\ 0 & \text{otherwise,} \end{cases} \tag{9.3}$$

where, by definition, $\widetilde{\pi}$ is the **extension** of π to a permutation of $\{0, 1, \ldots, n+1\}$ by $\widetilde{\sigma}(0) := 0$ and $\widetilde{\sigma}(n+1) := n+1$. In particular $[M\mathbf{p}]_1 = p_1$, $[M\mathbf{p}]_{n+1} = p_{n+1}$. It is easy to check that

$$\Delta_\pi(\mathbf{p}, \mathbf{p}') = \prod_{j=1}^{n+1} \delta\Big(p'_j - [M\mathbf{p}]_j \Big) \,, \tag{9.4}$$

in other words, each p'-momentum can be expressed as a linear combination of p-momenta, the matrix M encodes the corresponding coefficients, and these are all the relations among the p and p' momenta that are enforced by Δ_π. In particular, all p-momenta are independent.

The rule to express p'-momenta in terms of p-momenta is transparent in the graphical representation of the Feynman graph: the momentum p_j appears in those p'_i-momenta which fall into its "domain of dependence", i.e. the section between the image of the two endpoints of p_j, and the sign depends on the ordering of these images (Fig. 22). Notice that the roles of $\mathbf{p}$ and $\mathbf{p}'$ are symmetric, we could have expressed the p-momenta in terms of p'-momenta as well. It follows from this symmetry that $M(\pi^{-1}) = [M(\pi)]^{-1}$ and

$$\Delta_\pi(\mathbf{p}, \mathbf{p}') = \prod_{j=1}^{n+1} \delta\Big(p_j - [M^{-1}\mathbf{p}']_j \Big)$$

also holds.

A little linear algebra and combinatorics reveals that M is actually a **totally unimodular** matrix, which means that all its subdeterminants are 0 or ± 1. This will mean that the Jacobians of the necessary changes of variables are always controlled.

The following definition is crucial. It establishes the necessary concepts to measure the complexity of a permutation.

Definition 9.1 (Valley, peak, slope and ladder)**.** *Given a permutation $\pi \in \mathcal{S}_n$ let $\widetilde{\sigma}$ be its extension. A point $(j, \pi(j))$, $j \in I_n = \{1, 2, \ldots, n\}$, on the graph of π is called* **peak** *if $\pi(j) < \min\{\widetilde{\pi}(j-1), \widetilde{\pi}(j+1)\}$, it is called* **valley** *if $\pi(j) > \max\{\widetilde{\pi}(j-1), \widetilde{\pi}(j+1)\}$. Furthermore, if $\pi(j) - 1 \in \{\widetilde{\pi}(j-1), \widetilde{\pi}(j+1)\}$ and $(j, \pi(j))$ is not a valley, then the point $(j, \pi(j))$, $j \in I_n$, is called* **ladder**. *Finally, a point $(j, \pi(j))$, $j \in I_n$, on the graph of π is called* **slope** *if it is not a peak, valley or ladder.*

Let $I = \{1, 2, \ldots, n+1\}$ denote the set of row indices of M. This set is partitioned into five disjoint subsets, $I = I_p \cup I_v \cup I_\ell \cup I_s \cup I_{last}$, such that $I_{last} := \{n+1\}$ is the last index, and $i \in I_p, I_v, I_\ell$ or I_s depending on whether $(\pi^{-1}(i), i)$ is a peak, valley, ladder or slope, respectively. The cardinalities of these sets are denoted by $p := |I_p|$, $v := |I_v|$, $\ell := |I_\ell|$ and $s := |I_s|$. The dependence on π is indicated as $p = p(\pi)$ etc. if necessary. We define the **degree** *of the permutation π as*

$$\deg(\pi) := d(\pi) := n - \ell(\pi) \ . \tag{9.5}$$

Remarks: (i) The terminology of peak, valley, slope, ladder comes from the graph of the permutation $\widetilde{\pi}$ drawn in a coordinate system where the axis of the dependent variable, $\pi(j)$, is oriented downward (see Fig. 23). It immediately follows from the definition of the extension $\widetilde{\pi}$ that the number of peaks and valleys is the same:

$$p = v.$$

By the partitioning of I, we also have

$$p + v + \ell + s + 1 = n + 1.$$

(ii) The nonzero entries in the matrix $M(\sigma)$ follow the same geometric pattern as the graph: each downward segment of the graph corresponds to a column with a few consecutive 1's, upward segments correspond to columns with (-1)'s. These blocks of nonzero entries in each column will be called the **tower** of that column. On Fig. 23 we also pictured the towers of $M(\pi)$ as rectangles.

(iii) Because our choice of orientation of the vertical axis follows the convention of labelling rows of a matrix, a peak is a local minimum of $j \to \pi(j)$. We fix the convention that the notions "higher" or "lower" for objects related to the vertical axis (e.g. row indices) always refer to the graphical picture. In particular the "bottom" or the "lowest element" of a tower is located in the row with the highest index.

Also, a point on the graph of the function $j \to \pi(j)$ is traditionally denoted by $(j, \pi(j))$, where the first coordinate j runs on the horizontal axis, while in the labelling of the (i, j)–matrix element M_{ij} of a matrix M the first coordinate i labels rows, i.e. it runs vertically. To avoid confusion, we will always specify whether a double index (i, j) refers to a point on the graph of π or a matrix element.

(iv) We note that for the special case of the identity permutation $\pi = id = id_n$ we have $I_p = I_s = I_v = \emptyset$, and $I_\ell = \{1, 2, \ldots, n\}$. In particular, $\deg(id) = 0$ and $\deg(\pi) \geq 2$ for any other permutation $\pi \neq id$.

(v) Note that Definition 9.5 slightly differs from the preliminary definition of the degree given in (8.19). Not every index participating in a ladder are defined to be a ladder index; the

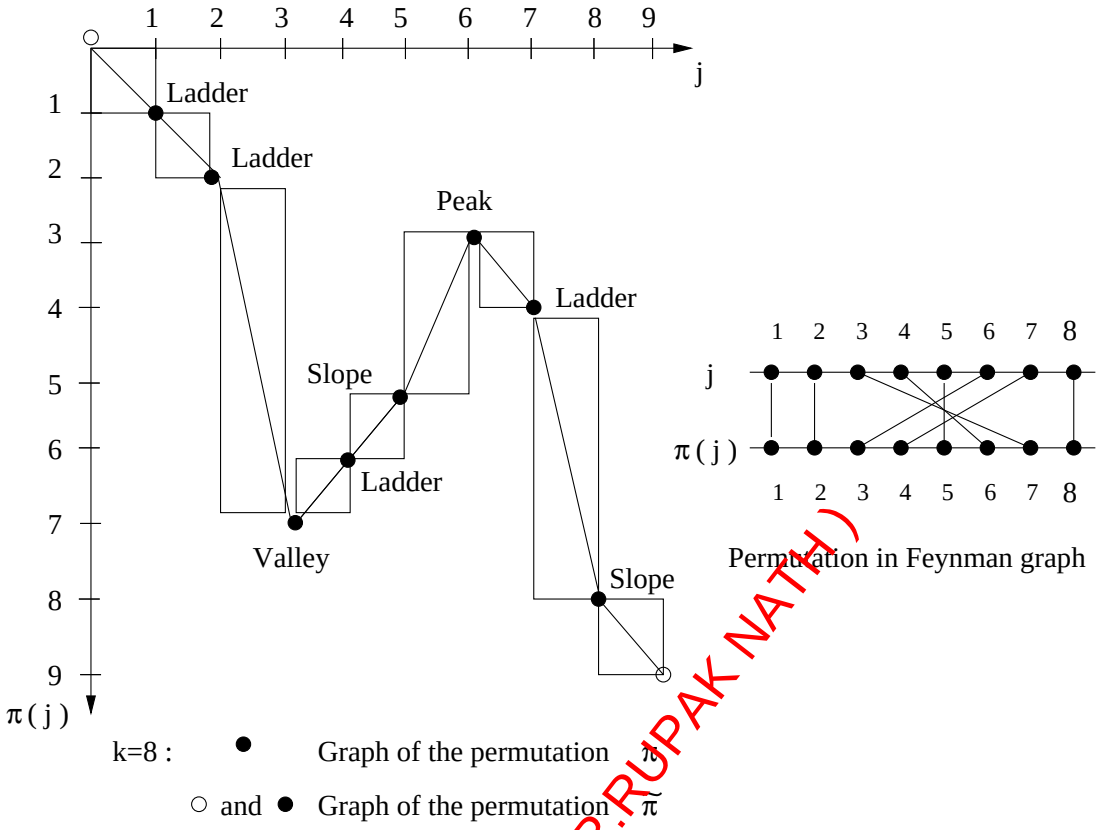

Figure 23: Graph of a permutation with the towers

top index of the ladder is excluded (to avoid overcounting) and the bottom index of a ladder is also excluded if it is a valley, since the valleys will play a special role. The precise definition of the degree given in Definition 9.5 is not canonical, other variants are also possible; the essential point is that long ladders should reduce the degree.

An example is shown on Fig. 23 with $n = 8$. The matrix corresponding to the permutation on this figure is the following (zero entries are left empty)

$$M(\pi) := \begin{pmatrix} 1 & & & & & & & \\ & 1 & & & & & & \\ & & 1 & & & & & \\ & & 1 & & & -1 & 1 & \\ & & 1 & & & -1 & & 1 \\ & & 1 & & -1 & & & 1 \\ & & 1 & -1 & & & & 1 \\ & & & & & & & 1 \\ & & & & & & 1 & \end{pmatrix} \begin{matrix} 1 & \ell \\ 2 & \ell \\ 3 & p \\ 4 & \ell \\ 5 & s \\ 6 & \ell \\ 7 & v \\ 8 & s \\ 9 & (last) \end{matrix} \tag{9.6}$$

The numbers on the right indicate the column indices and the letters show whether it is peak/slope/valley/ladder or last. In this case $I_p = \{3\}$, $I_v = \{7\}$, $I_s = \{5, 8\}$. $I_\ell = \{1, 2, 4, 6\}$, $I_{last} = \{9\}$ and $\deg(\sigma) = 4$.

9.2 Failed attempts to integrate out Feynman diagrams

To estimate the value of a Feynman graph, based upon (9.1) and (9.4), our task is to successively integrate out all p_j's in the multiple integral of the form

$$Q(M) := \lambda^{2n} \int \mathrm{d}\alpha \mathrm{d}\beta \int \mathrm{d}\mathbf{p} \prod_{i=1}^{n+1} \frac{1}{|\alpha - \omega(p_i) - i\eta|} \frac{1}{\left|\beta - \omega\big(\sum_{j=1}^{n+1} M_{ij} p_j\big) + i\eta\right|}$$

As usual, the unspecified domains of integrations for the α and β variables is $\mathbb{R}$, and for the $\mathrm{d}\mathbf{p}$ momentum variables is $(\mathbf{T}^d)^{n+1}$. As one p_j is integrated out, the matrix M needs to be updated and effective algorithm is needed to keep track of these changes.

Why is it hard to compute or estimate such a quite explicit integral? The problem is that there is nowhere to start: each integration variable p_j may appear in many denominators: apparently there is no "easy" integration variable at the beginning.

As a **first attempt**, we can try to decouple the interdependent denominators by Schwarz or Hölder inequalities. It turns out that they cannot be used effectively, since by their application we will lose the non-overlapping effects imposed by the crossing. If one tries to decouple these integrals trivially, one obtains the ladder and gains nothing. In contrast to a complicated crossing diagram, the ladder is easily computable because $p_i = p_i'$ means that the integrals decouple:

$$\begin{aligned} Q(M) \leq \lambda^{2n} \int \mathrm{d}\alpha \mathrm{d}\beta \int \mathrm{d}\mathbf{p} & \frac{1}{|\alpha - \omega(p_{n+1}) - i\eta|} \frac{1}{\left|\beta - \omega(p_{n+1}) + i\eta\right|} \\ & \times \left[\prod_{i=1}^{n} \frac{1}{|\alpha - \omega(p_i) - i\eta|^2} + \frac{1}{\left|\beta - \omega\big(\sum_{j=1}^{n} \widetilde{M}_{ij} p_j\big) + i\eta\right|^2} \right], \end{aligned} \tag{9.7}$$

where $\widetilde{M}$ denotes the $n \times n$ upper minor of M and we used that M and $\widetilde{M}$ differ only with an entry 1 in the diagonal, so in particular $\sum_{j=1}^{n+1} M_{n+1,j} p_j = p_{n+1}$ and $\det(M) = \det(\widetilde{M})$. Changing variables in the second term ($p_i' = \sum_{j=1}^{n} \widetilde{M}_{ij} p_j$) and using that the Jacobian is one ($\det(M) = \pm 1$), we see that the first and second terms in the parenthesis are exactly the same. Thus

$$Q(M) \leq 2\lambda^{2n} \int \mathrm{d}\alpha \mathrm{d}\beta \int \mathrm{d}\mathbf{p} \frac{1}{|\alpha - \omega(p_{n+1}) - i\eta|} \frac{1}{\left|\beta - \omega(p_{n+1}) + i\eta\right|} \prod_{i=1}^{n} \frac{1}{|\alpha - \omega(p_i) - i\eta|^2},$$

and we can succesively integrate out the momenta $p_1, p_2, \ldots p_n$ and then finally α, β. Using the inequalities (8.23) and (8.25), we obtain (with the choice $\eta = 1/t$)

$$|Val(\pi)| \leq e^{2\eta t} Q(M) \leq C\lambda^{2n} \lambda^{-2n} (|\log \eta|^2) = C(|\log \eta|^2)$$

i.e. essentially the same estimate as the value of the ladder (8.27).

As a **second attempt**, we can trivially estimate all (but one) β-denominators, by their L^∞ norm

$$\left| \frac{1}{\beta - \omega(\ldots) + i\eta} \right| \leq \frac{1}{\lambda^2 |\mathrm{Im}\omega|} \sim \lambda^{-2} \tag{9.8}$$

(see (8.18) and the remark afterward on its limitations), and then integrate out all p_j one by one by using the L^1-bound (8.26)

$$\int \frac{\mathrm{d}p}{|\alpha - \omega(p) + i\eta|} = O(|\log \eta|).$$

This gives

$$|Val(\pi)| \leq O(|\log \eta|^{n+2}),$$

i.e. it is essentially order 1 with logarithmic corrections. Note that this second method gives a worse exponent for the logarithm than the first one, nevertheless this method will be the good starting point.

The second attempt is a standard integration procedure in diagrammatic perturbation theory. The basic situation is that given a large graph with momenta assigned to the edges, each edge carries a propagator, i.e. a function depending in this momenta, and momenta are subject to momentum conservation (Kirchoff law) at each vertex. The task is to integrate out all momenta. The idea is to use L^∞-bounds for the propagator on certain edges in the graph to decouple the rest and use independent L^1-bounds for the remaining integrations. Typically one constructs a spanning tree in the graph, then each additional edge creates a loop. The Kirchoff delta functions amount to expressing all tree momenta in terms of loop momenta in a trivial way. Then one uses L^∞-bound on "tree"-propagators to free up all Kirchoff delta functions, integrate out the "tree"-variables freely and then use L^1-bound on the "loop"-propagators that are now independent.

This procedure can be used to obtain a rough bound on values of very general Feynman diagrams subject to Kirchoff laws. We will formalize this procedure in Appendix B. Since typically L^1 and L^∞ bounds on the propagators scale with a different power of the key parameter of the problem (in our case λ), we will call this method *the power counting estimate*.

In our case simple graph theoretical counting shows that

$$\text{Number of tree momenta} = \text{number of loop momenta} = n + 1.$$

In fact, after identifying the paired vertices in the Feynman diagram (after closing the two long horizontal lines at the two ends), we obtain a new graph where one can easily see that the edges in the lower horizontal line, i.e. the edges corresponding p'_j-momenta, form a spanning tree and the edges of all p_j-variables form loops. Thus the delta functions in (9.4) represent the way how the tree momenta are expressed in terms of the loop momenta.

Since each L^∞-bound on "tree"-propagators costs λ^{-2} by (9.8), the total estimate would be of order

$$\lambda^{2n}\lambda^{-2(n+1)}$$

with actually logarithmic factors. But due to the additional α, β integrations, one L^∞-bound can be saved (modulo logarithm) by using (8.25). So the total estimate is, in fact,

$$\lambda^{2n}\lambda^{-2n} = O(1)$$

modulo logarithmic factors, i.e. the same size as for the ladder diagrams. The conclusion is that even in the second attempt, with the systematic power counting estimate, we did not gain from the crossing structure of the permutation either.

9.3 The new algorithm to integrate out Feynman diagrams

In our new algorithm, we use the L^∞-bound on the propagators for a carefully selected subset of the "tree"-variables: the ones that lie "above the peak". The selection is shown in Fig. 24 for a concrete example. Notice that the segments in the horizontal axis correspond to $p_1, p_2, \dots, p_{n+1}$, i.e. the loop momenta, and the segments in the vertical axis correspond to the tree momenta.

We will first explain the main idea, then we will work out a concrete example to illustrate the algorithm. After drawing the graph of the permutation, a structure of valleys and peaks emerges. By carefully examining the relation between this graph and the matrix M, one notices that if one estimates only those "tree"-propagators that lie above a peak by the trivial L^∞-bound, then all the remaining propagators can be successively integrated out by selecting the integration variables p_j in an appropriate order (dictated by the graph). Each integration involves no more than two propagators at one time, hence it can be estimated by elementary calculus (using estimates collected in Appendix A).

In fact, this is correct if there are no ladder indices; but momenta corresponding to ladder indices can be integrated out locally (see later). As it turns out, ladder indices are neutral; their integration yields an $O(1)$ factor. The gain comes from non-ladder indices, and this will justify the definition of the degree (9.5).

More precisely, after an easy bookkeeping we notice that in this way we gain roughly as many λ^2 factors as many slopes and valleys we have (no gain from peaks or ladders). Since the number of peaks and valleys are the same,

$$p = v,$$

and the peaks, valleys, slopes and ladders altogether sum up to n,

$$p + v + \ell + s = n$$

(see Remark (i) after Definition 9.1), we see that

$$v + s = n - p - \ell \geq \frac{1}{2}(n - \ell) = \frac{1}{2}d(\pi),$$

since $n - 2p - \ell = n - p - v - \ell = s \geq 0$. Thus we gain at least $\lambda^{d(\pi)}$. This would prove (8.20) with $\kappa = 1$. After various technical estimates that were neglected here, the actual value of κ is reduced, but it still remains positive and this completes the proof of Lemma 8.6. □

9.3.1 An example without ladder

As an example, we will show the integration procedure for the graph on Fig. 24. This is a permutation which has no ladder index for simplicity, we will comment on ladder indices afterwards. The total integral is (for simplicity, we neglected the $\pm i\eta$ regularization in the

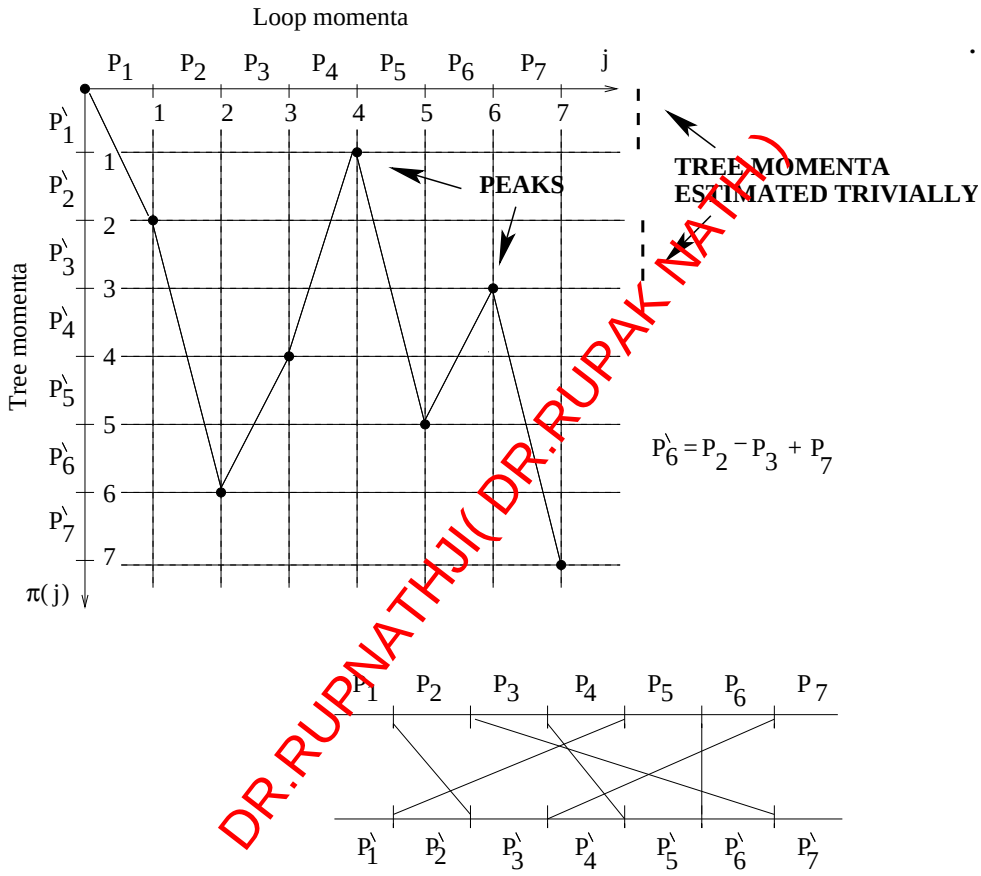

Figure 24: Valleys and peaks determine the order of integration

formulas below to highlight the structure better)

$$\lambda^{12}\int \mathrm{d}\alpha\mathrm{d}\beta\mathrm{d}\mathbf{p}\frac{1}{|\alpha-\omega(p_1)|}\frac{1}{|\alpha-\omega(p_2)|}\frac{1}{|\alpha-\omega(p_3)|}\frac{1}{|\alpha-\omega(p_4)|}\frac{1}{|\alpha-\omega(p_5)|}\frac{1}{|\alpha-\omega(p_6)|}\frac{1}{|\alpha-\omega(p_7)|}$$
$$\times\frac{1}{|\beta-\omega(p_1)|}\frac{1}{|\beta-\omega(p_1-p_4+p_5)|}\frac{1}{|\beta-\omega(p_2-p_4+p_5)|}\frac{1}{|\beta-\omega(p_2-p_4+p_5-p_6+p_7)|}$$
$$\times\frac{1}{|\beta-\omega(p_2-p_3+p_5-p_6+p_7)|}\frac{1}{|\beta-\omega(p_2-p_3+p_7)|}\frac{1}{|\beta-\omega(p_7)|}. \tag{9.9}$$

Notice that every variable appears at least in three denominators, for a graph of order n, typically every momentum appears in $(const)n$ different denominators, so there is no way to start the integration procedure (an integral with many almost singular denominators is impossible to estimate directly with a sufficient precision).

Now we estimate those β-factors by L^∞-norm (i.e. by λ^{-2} according to (9.8)) whose corresponding primed momenta lied right above a peak; in our case these are the factors

$$\frac{1}{|\beta-\omega(p_1')|}\frac{1}{|\beta-\omega(p_3')|}$$

i.e. the first and the third β-factor in (9.9). After this estimate, they will be removed from the integrand. We will lose a factor $(\lambda^{-2})^p=\lambda^{-4}$ recalling that $p=2$ is the number of peaks. We are left with the integral

$$\lambda^{12}(\lambda^{-2})^p\int \mathrm{d}\alpha\mathrm{d}\beta\mathrm{d}\mathbf{p}\frac{1}{|\alpha-\omega(p_1)|}\frac{1}{|\alpha-\omega(p_2)|}\frac{1}{|\alpha-\omega(p_3)|}\frac{1}{|\alpha-\omega(p_4)|}\frac{1}{|\alpha-\omega(p_5)|}\frac{1}{|\alpha-\omega(p_6)|}$$
$$\times\frac{1}{|\alpha-\omega(p_7)|}\frac{1}{|\beta-\omega(p_1-p_4+p_5)|}\frac{1}{|\beta-\omega(p_2-p_4+p_5-p_6+p_7)|}$$
$$\times\frac{1}{|\beta-\omega(p_2-p_3+p_5-p_6+p_7)|}\frac{1}{|\beta-\omega(p_2-p_3+p_7)|}\frac{1}{|\beta-\omega(p_7)|}. \tag{9.10}$$

Now the remaining factors can be integrated out by using the following generalized version of (8.29) (see Appendix A)

$$\sup_{\alpha,\beta}\int\frac{\mathrm{d}p}{|\alpha-\omega(p)+i\eta|}\frac{1}{|\beta-\omega(p+u)-i\eta|}\le\frac{|\log\eta|}{|u|+\lambda^2}. \tag{9.11}$$

Suppose we can forget about the possible point singularity, i.e. neglect the case when $|u|\ll 1$. Then the integral (9.11) is $O(1)$ modulo an irrelevant log factor. Then the successive integration is done according to graph: we will get rid of the factors $|\beta-\omega(p_1')|^{-1}$, $|\beta-\omega(p_2')|^{-1}$, $|\beta-\omega(p_3')|^{-1}$, etc., in this order. The factors with p_1' and p_3' have already been removed, so the first nontrivial integration will eliminate

$$\frac{1}{|\beta-\omega(p_2')|}=\frac{1}{|\beta-\omega(p_1-p_4+p_5)|}. \tag{9.12}$$

Which integration variable to use? Notice that the point (1,2) was not a peak, that means that there is a momentum (in this case p_1) such that p_2' is the primed momenta with the

largest index that still depends on p_1 (the "tower" of p_1 ends at p_2'). This means that among all the remaining β-factors no other β-factor, except for (9.12), involves p_1! Thus only two factors involve p_1 and not more, so we can integrate out the p_1 variable by using (9.11)

$$\int dp_1 \frac{1}{|\alpha-\omega(p_1)|}\frac{1}{|\beta-\omega(p_1-p_4+p_5)|}=O(1)$$

(modulo logs and modulo the point singularity problem).

In the next step we are then left with

$$\lambda^{12}(\lambda^{-2})^p\int \mathrm{d}\alpha\mathrm{d}\beta\mathrm{d}\mathbf{p}\frac{1}{|\alpha-\omega(p_2)|}\frac{1}{|\alpha-\omega(p_3)|}\frac{1}{|\alpha-\omega(p_4)|}\frac{1}{|\alpha-\omega(p_5)|}\frac{1}{|\alpha-\omega(p_6)|}$$
$$\times\frac{1}{|\alpha-\omega(p_7)|}\frac{1}{|\beta-\omega(p_2-p_4+p_5-p_6+p_7)|}$$
$$\times\frac{1}{|\beta-\omega(p_2-p_3+p_5-p_6+p_7)|}\frac{1}{|\beta-\omega(p_2-p_3+p_7)|}\frac{1}{|\beta-\omega(p_7)|}.$$

Since we have already taken care of the β-denominators with p_1', p_2', the next one would be the β-denominator with p_3', but this was estimated trivially (and removed) at the beginning. So the next one to consider is

$$\frac{1}{|\beta-\omega(p_4')|}=\frac{1}{|\beta-\omega(p_2-p_4+p_5-p_6+p_7)|}.$$

Since p_4' is not above a peak, there is a p-momentum whose tower has the lowest point at the level 4, namely p_4. From the graph we thus conclude that p_4 appears only in this β-factor (and in one α-factor), so again it can be integrated out:

$$\int \mathrm{d}p_4\frac{1}{|\alpha-\omega(p_4)|}\frac{1}{|\beta-\omega(p_2-p_4+p_5-p_6+p_7)|}\le O(1)$$

(modulo log's and point singularity).

We have then

$$\lambda^{12}(\lambda^{-2})^p\int \mathrm{d}\alpha\mathrm{d}\beta\mathrm{d}\mathbf{p}\frac{1}{|\alpha-\omega(p_2)|}\frac{1}{|\alpha-\omega(p_3)|}\frac{1}{|\alpha-\omega(p_5)|}\frac{1}{|\alpha-\omega(p_6)|}$$
$$\times\frac{1}{|\alpha-\omega(p_7)|}\frac{1}{|\beta-\omega(p_2-p_3+p_5-p_6+p_7)|}\frac{1}{|\beta-\omega(p_2-p_3+p_7)|}\frac{1}{|\beta-\omega(p_7)|}.$$

Next,

$$\frac{1}{|\beta-\omega(p_5')|}=\frac{1}{|\beta-\omega(p_2-p_3+p_5-p_6+p_7)|}$$

includes even two variables (namely p_5, p_6) that do not appear in any other β-denominators (because the towers of p_5 and p_6 end at the level 5, in other words because right below the row of p_5' there is a valley). We can freely choose which one to integrate out, say we choose p_5, and perform

$$\int \mathrm{d}p_5\frac{1}{|\alpha-\omega(p_5)|}\frac{1}{|\beta-\omega(p_2-p_3+p_5-p_6+p_7)|}\le O(1).$$

We are left with

$$\lambda^{12}(\lambda^{-2})^p \int \mathrm{d}\alpha \mathrm{d}\beta \mathrm{d}\mathbf{p} \frac{1}{|\alpha-\omega(p_2)|}\frac{1}{|\alpha-\omega(p_3)|}\frac{1}{|\alpha-\omega(p_6)|} \times \frac{1}{|\alpha-\omega(p_7)|}\frac{1}{|\beta-\omega(p_2-p_3+p_7)|}\frac{1}{|\beta-\omega(p_7)|}. \tag{9.13}$$

Finally

$$\frac{1}{|\beta-\omega(p_6')|} = \frac{1}{|\beta-\omega(p_2-p_3+p_7)|}$$

can be integrated out either by p_2 or p_3 (both towers end at the level of p_6'), e.g.

$$\int dp_2 \frac{1}{|\alpha-\omega(p_2)|}\frac{1}{|\beta-\omega(p_2-p_3+p_7)|} \le O(1).$$

The last β-factor is eliminated by the β integration by (8.25) and then in the remaining integral,

$$\lambda^{12}(\lambda^{-2})^p \int \mathrm{d}\alpha \mathrm{d}\mathbf{p} \frac{1}{|\alpha-\omega(p_3)|}\frac{1}{|\alpha-\omega(p_6)|}\frac{1}{|\alpha-\omega(p_7)|},$$

one can integrate out each remaining momenta one by one. We have thus shown that the value of this permutation is

$$\mathrm{Val}(\pi) \le \lambda^{12}(\lambda^{-2})^p = \lambda^8$$

modulo log's and point singularities.

In general, the above procedure gives

$$\lambda^{2n-2p} \le \lambda^n = \lambda^{d(\pi)}$$

if there are no ladders, $\ell = 0$.

9.3.2 General algorithm including ladder indices and other fine points

Finally, we show how to deal with ladder indices. The idea is that first one has to integrate out the consecutive ladder indices after taking a trivial Schwarz inequality to decouple the α and β-factors of the consecutive ladder indices and then proceed similarly to (9.7). It is important that only propagators with ladder indices will be Schwarzed, for the rest of the integrand we will use the successive integration procedure explained in the previous section. In this way the ladder indices remain neutral for the final bookkeeping, in fact, the integral

$$\int \frac{\mathrm{d}p}{|\alpha-\omega(p)-i\eta|^2} \sim \lambda^{-2}$$

exactly compensates the λ^2 prefactor carried by the ladder index. Actually, one needs to take care that not only the λ-powers but even the constants cancel each other as well, since an error C^n would not be affordable when n, being the typical number of the collisions, is $\lambda^{-\kappa}$. Therefore the above bound will be improved to

$$\int \frac{\lambda^2}{|\alpha-\omega(p)-i\eta|^2}\mathrm{d}p = 1 + O(\lambda^{1-12\kappa}), \tag{9.14}$$

(see (A.9)), and it is this point where the careful choice of the remormalized propagator $\omega(p)$ plays a crucial role. This is also one error estimate which further restricts the value of κ. After integrating out the ladder indices, we perform the successive estimates done in Section 9.3.1.

There were two main issues have been swept under the rug (among several other less important ones...). First, there are several additional log factors floating around. In fact, they can be treated generously, since with this procedure we gain a $\lambda^{(const.)d(\pi)}$ factor and the number of log factors is also comparable with $d(\pi)$. However, here the key point again is that by the ladder integration in (9.14) we do not lose any log factor; even not a constant factor!

The second complication, the issue of the point singularities, is much more serious and this accounts for the serious reduction of the value of κ in the final theorem. In higher dimensions, $d \geq 3$, one point singularity of the form $(|p| + \eta)^{-1}$ is integrable, but it may happen that the *same* point singularity arises from different integrations of the form (9.11) along our algorithm. This would yield a high multiplicity point singularity whose integral is large.

First, we indicate with an example that overlapping point singularities indeed do occur; they certainly would occur in the ladders, had we not integrated out the ladders separately.

The structure of the integral for a set of consecutive ladder indices, $\{k, k+1, k+2, \dots k+m\}$ is as follows:

$$\Omega = \int \mathrm{d}p_k \mathrm{d}p_{k+1} \dots \mathrm{d}p_{k+m} \frac{1}{|\alpha - \omega(p_k)|} \frac{1}{|\alpha - \omega(p_{k+1})|} \cdots \frac{1}{|\alpha - \omega(p_{k+m})|} \times \frac{1}{|\beta - \omega(p_k + u)|} \frac{1}{|\beta - \omega(p_{k+1} + u)|} \cdots \frac{1}{|\beta - \omega(p_{k+m} + u)|}. \tag{9.15}$$

Here we used that if the consecutive ladders are the points $(k, s), (k+1, s+1), (k+2, s+2), \dots$, then the corresponding momenta are related as

$$p_k - p'_s = p_{k+1} - p'_{s+1} = p_{k+2} - p'_{s+2} = \dots$$

i.e. one can write $p'_{s+i} = p_{k+i} + u$ with a fixed vector u (that depends on all other momenta but not on $p_k, p_{k+1}, \dots p_{k+m}$, e.g. $u = p_2 - p_4 + p_9$).

Using (9.11) successively, we obtain

$$\Omega \leq \Big(\frac{1}{|u| + \lambda^2} \Big)^{m+1},$$

i.e. the **same** point singularity arises from each integration. We call this phenomenon *accumulation of point singularities*. Since u is a linear combination of other momenta, that need to be integrated out, at some point we would face with

$$\int \Big(\frac{1}{|u| + \lambda^2} \Big)^{m+1} \mathrm{d}u \sim \Big(\lambda^{-2} \Big)^{m-2}. \tag{9.16}$$

Since $m + 1$ consecutive ladder indices carry a factor $\lambda^{2(m+1)}$, we see that from a consecutive ladder sequence we might gain only λ^6, irrespective of the length of the sequence. It turns out that even this calculation is a bit too optimistic; the formula (9.16) is only rough caricature,

also propagators depending on u are present in the u integration. In reality, eventually, the ladder indices are integrated out to $O(1)$.

Based upon this example, one may think that the accumulation of point singularities occurs only in the presence of the consecutive ladders; once they are taken care of separately (and not gaining but also not losing from them), the remaining graph can be integated out without any accumulation of point singularities. We conjecture that this is indeed the case, but we could not quite prove this. Instead, we estimate several more β-factors (propagators with "tree momenta") by the trivial L^∞-bound to surely exclude this scenario. They are carefully chosen (see the concept of "uncovered slope indices" in Definition 10.3 of [23]) so that after their trivial removal, along the successive integration algorithm for the remaining propagators, indeed no accumulation of point singularity occurs.

This completes the sketch of the proof of estimating the non-repetition Feynman diagrams, i.e. the proof of Theorem 8.2. □

10 Feynman graphs with repetitions

In this short section we indicate how to prove Theorem 8.1. Recall that ψ_s^{err} contains terms with many ($n \gg \lambda^{-\kappa}$) potentials and terms with (not immediate) recollisions.

The estimate of the terms with a large number of collisions is relatively easy; these are still non-repetition graphs, so the integration procedure from Section 9 applies. Permutations with degree $d(\pi) \geq C/\kappa$ have a much smaller contribution than the required $o(t^{-2})$ error. Permutations with low complexity contain macroscopically long ladders (i.e. ladders with length cn with some positive constant c) and these can be computed quite precisely. The precise calculation reveals a factor $1/(cn)!$ due to the time-ordered integration (see discussion around (7.12)) and the value of such graphs can be estimated by

$$\frac{(\lambda^2 t)^n}{(cn)!} e^{-c'\lambda^2 t}$$

The combinatorics of such low complexity graphs is at most $n^{C/\kappa}$, so their total contribution is negligible if $n \gg \lambda^2 t \sim \lambda^{-\kappa}$.

The repetition terms from ψ_s^{err} require a much more laborous treatment. The different repetition patterns are shown on Fig. 25. When one of the repetitions shows up (apart from immediate repetition that were renormalized), we stop the expansion to reduce complications (see (8.1)). Actually the stopping rule is a bit more involved, because the repetition pattern has to collect sufficient "repetitions" to render that term sufficiently small even after paying the price for the unitary estimate, but we will no go into these details. Finally, each term we compute by "bare hand", after applying a certain graph surgery to reduce the number of different cases. One example of this reduction is shown on Fig. 26 while Fig. 27 shows the result of an explicit estimate. The explicit estimates rely on the fact that the repetition imposes a certain momentum restriction that reduces the volume of the maximal overlap of singularities, reminiscent to the mechanism behind the crossing estimate, Section 8.6. Unfortunately, even after the graph surgery reduction, still a considerably number of similar but not identical cases have to be estimated on a case by case basis.

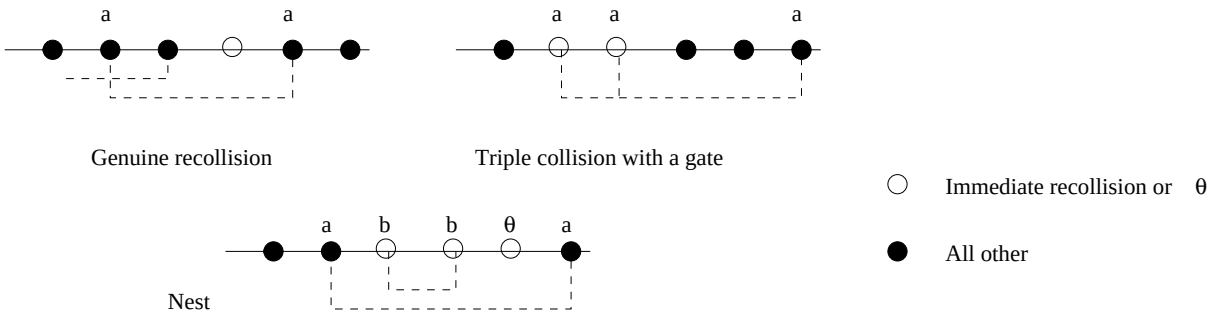

Figure 25: Various repetition patterns

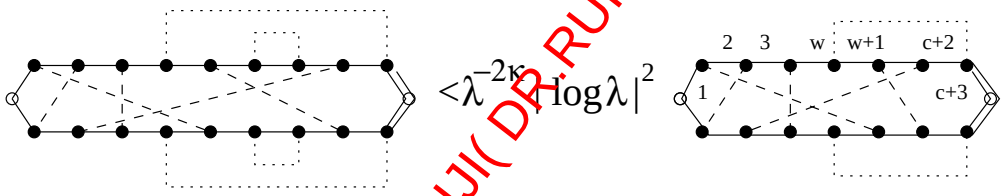

Figure 26: Removal of a gate

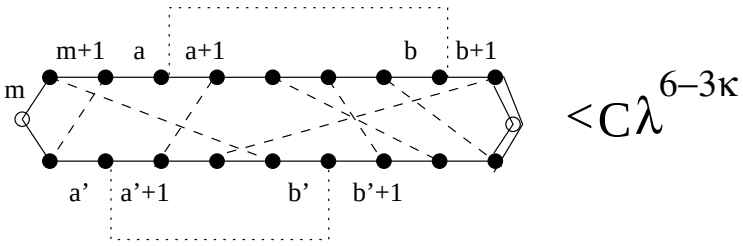

Figure 27: Two sided recollision

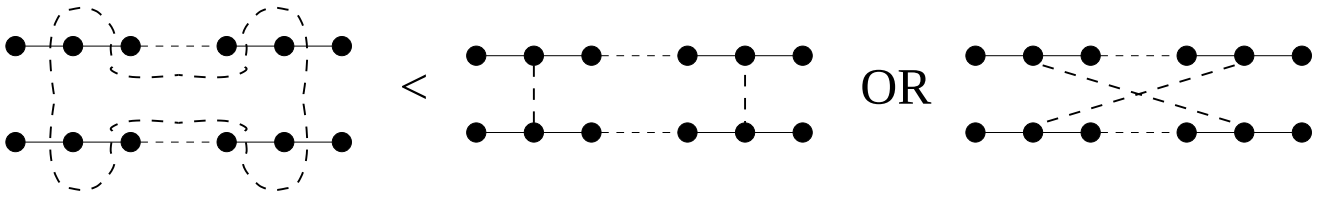

Figure 28: Breaking up lumps

10.1 Higher order pairings: the lumps

We also mention how to deal with the higher order pairings that we postponed from Section 8.2. Recall that the stopping rule monitored whether we have a (non-immediate) repetition in the collision sequence $A = (\alpha_1, \alpha_2, \ldots, \alpha_n)$. This procedure has a side effect when computing the expectation:

$$\mathbf{E} \prod \overline{\widehat{V}_{\alpha_j}(p_{j+1} - p_j)} \widehat{V}_{\alpha_j}(q_{j+1} - q_j) = \sum_{\alpha_\ell \neq \alpha_k} \prod_j e^{i\alpha_j [p_{j+1} - p_j - (q_{j+1} - q_j)]}$$

The non-repetition restriction $\alpha_\ell \neq \alpha_k$ destroys the precise delta function $\sum_\alpha e^{i\alpha p} = \delta(p)$. This requires to employ the following Connected Graph Formula that expresses the restricted sum as a linear combination of momentum delta functions:

Lemma 10.1 (Connected Graph Formula). *Let $\mathcal{A}_n$ be the set of partitions of $\{1, \ldots, n\}$. There exist explicit coefficients $c(k)$ such that*

$$\sum_{\alpha_\ell \neq \alpha_k} e^{iq_j \alpha_j} = \sum_{\mathbf{A} \in \mathcal{A}_n} \prod_\nu c(|A_\nu|) \delta\Big(\sum_{\ell \in A_\nu} q_\ell \Big) \qquad \mathbf{A} = (A_1, A_2, \ldots),$$

where the summation is over all partitions $\mathbf{A}$ of the set $\{1, \ldots, n\}$, and $A_1, A_2, \ldots,$ denote the elements of a fixed partition, in particular they are disjoint sets and $A_1 \cup A_2 \cup \ldots = \{1, \ldots, n\}$.

The appearence of nontrivial partitioning sets means that instead of pairs we really have to consider "hyperpairs", i.e. subsets (called *lumps*) of size more than 2.

Lumps have large combinatorics, but the value of the corresponding Feynman graph is small since in a large lump many more obstacle indices must coincide than in a pairing. Nevertheless, their treatment would require setting up a separate notation and run the algorithm of Section 9 for general lumps instead of pairs.

To avoid these complications and repeating a similar proof, we invented a graph surgery that reduces lumps to pairs. The idea is that lumps are artificially broken up into pairs (i.e. permutations) and the corresponding Feynman graphs can be compared. The break-up is not unique, e.g. Fig. 28 shows two possible break-ups of a lump of four elements into two pairs.

There are many break-ups possible, but the key idea is that we choose the break-up that gives the biggest $d(\pi)$. The following combinatorial statement shows a lower estimate on the maximal degree, demonstrating that every lump can be broken up into an appropriate pairing with relatively high degree.

Proposition 10.2. *Let* $\mathbf{B} = \{B_1, B_2, \ldots\}$ *be a partition of vertices.*

$$s(\mathbf{B}) := \frac{1}{2} \sum \{\, |B_j| \;:\; |B_j| \geq 4\}$$

Then there exists a permutation π*, compatible with* $\mathbf{B}$ *such that* $d(\pi) \geq \frac{1}{2} s(\mathbf{B})$*.*

Using this lemma we reduce the estimate of hyperpairs to the previous case involving only pairs. This completes the very sketchy ideas of the proof of Theorem 8.1. □

11 Computation of the main term

In this section we sketch the computation of the main term, i.e. the proof of Theorem 8.3. It is an explicit but nontrivial computation. The calculation below is valid only in the discrete case as it uses the fact that for any incoming velocity u, the collision kernel $\sigma(u, v)$ is constant in the outgoing velocity, i.e. the new velocity is independent of the old one (this was the reason why the diffusion coefficient could be directly computed (see (6.9)) without computing the velocity autocorrelation function via the Green-Kubo formula (6.11) as in the continuous case.

The computation of the main term for the continuous case relates the diffusion coefficient to the underlying Boltzmann process. This procedure is conceptually more general but technically somewhat more involved (see Section 6 of [24]), so here we present only the simpler method for the discrete model.

The Fourier transform of the rescaled Wigner transform $W(X/\varepsilon, V)$ is given by

$$\widehat{W}_t(\varepsilon\xi, v) = \overline{\psi}_t\Big(v + \frac{\varepsilon\xi}{2}\Big)\widehat{\psi}_t\Big(v - \frac{\varepsilon\xi}{2}\Big),$$

where recall that $\varepsilon = \lambda^{2+\kappa/2}$ is the space rescaling and $t = \lambda^{-2-\kappa}T$.

We want to test the Wigner transform against a macroscopic observable, i.e. compute

$$\langle \mathcal{O}, \mathbf{E}\widehat{W}_t \rangle = \langle \mathcal{O}(\xi, v), \mathbf{E}\widehat{W}_t(\varepsilon\xi, v) \rangle = \int \mathrm{d}v\mathrm{d}\xi \; \mathcal{O}(\xi, v)\, \mathbf{E}\widehat{W}_t(\varepsilon\xi, v).$$

Recall from Lemma 7.1 that the Wigner transform enjoys the following continuity property:

$$\langle \mathcal{O}, \mathbf{E}\widehat{W}_\psi \rangle - \langle \mathcal{O}, \mathbf{E}\widehat{W}_\phi \rangle \leq C \sqrt{\mathbf{E}\|\psi\|^2 + \mathbf{E}\|\phi\|^2} \sqrt{\mathbf{E}\|\psi - \phi\|^2},$$

in particular, by using (8.3), it is sufficient to compute the Wigner transform of

$$\sum_{n=0}^{K} \psi_{n,t}^{nr}.$$

We can express this Wigner transform in terms of Feynman diagrams, exactly as for the L^2-calculation. The estimate (8.5) in our key Theorem 8.2 implies, that only the ladder diagrams

matter (after renormalization). Thus we have

$$\begin{aligned}\langle \mathcal{O}, \mathbf{E}\widehat{W}_t\rangle \approx \sum_{k\le K} \lambda^{2k} \int_{\mathbb{R}} \mathrm{d}\alpha\mathrm{d}\beta\, e^{it(\alpha-\beta)+2t\eta} \\ \times \int \mathrm{d}\xi\mathrm{d}v \mathcal{O}(\xi,v)\overline{R_\eta\Big(\alpha, v+\frac{\varepsilon\xi}{2}\Big)} R_\eta\Big(\beta, v-\frac{\varepsilon\xi}{2}\Big) \\ \times \prod_{j=2}^{k}\Big[\int \mathrm{d}v_j \overline{R_\eta\Big(\alpha, v_j+\frac{\varepsilon\xi}{2}\Big)} R_\eta\Big(\beta, v_j-\frac{\varepsilon\xi}{2}\Big)\Big] \\ \times \int \mathrm{d}v_1 \overline{R_\eta\Big(\alpha, v_1+\frac{\varepsilon\xi}{2}\Big)} R_\eta\Big(\beta, v_1-\frac{\varepsilon\xi}{2}\Big)\overline{\widehat{W}_0}(\varepsilon\xi, v_1)\ ,\end{aligned} \tag{11.1}$$

with the renormalized propagator

$$R_\eta(\alpha, v) := \frac{1}{\alpha - e(v) - \lambda^2\theta(v) + i\eta}$$

We perform each $\mathrm{d}v_j$ integral. The key technical lemma is the following:

Lemma 11.1. *Let $f(p) \in C^1(\mathbb{R}^d)$, $a := (\alpha+\beta)/2$, $\lambda^{2+4\kappa} \le \eta \le \lambda^{2+\kappa}$ and fix $r \in \mathbb{R}^d$ with $|r| \le \lambda^{2+\kappa/4}$. Then we have*

$$\begin{aligned}&\int \frac{\lambda^2 f(v)}{\Big(\alpha - e(v-r) - \lambda^2\overline{\theta}(v-r) - i\eta\Big)\Big(\beta - e(v+r) - \lambda^2\theta(v+r) + i\eta\Big)}\,\mathrm{d}v \\ &\quad = -2\pi i \int \frac{\lambda^2 f(v)\,\delta(e(v)-a)}{(\alpha-\beta) + 2(\nabla e)(v)\cdot r - 2i\lambda^2\mathcal{I}(a)}\,\mathrm{d}v + o(\lambda^{1/4})\end{aligned}$$

where

$$\mathcal{I}(a) := Im \int \frac{\mathrm{d}v}{a - e(v) - i0} = \int \delta(e(v)-a)\,\mathrm{d}v\ ,$$

in particular, $\mathcal{I}(e(p)) = Im\,\theta(p)$.

The proof of Lemma 11.1 relies on the following (approximate) identity

$$\begin{aligned}\frac{1}{(\alpha - \overline{g}(v-r) - i0)(\beta - g(v+r) + i0)} \approx \frac{1}{\alpha - \beta + g(v+r) - \overline{g}(v-r)} \\ \times\Big[\frac{1}{\beta - g(v) + i0} - \frac{1}{\alpha - \overline{g}(v) - i0}\Big]\end{aligned}$$

and on careful Taylor expansions. □

Accepting this lemma, we change variables $a = (\alpha+\beta)/2$, $b = (\alpha-\beta)/\lambda^2$ and choose $\eta \ll t^{-1}$ in (11.1). Then we get

$$\langle \mathcal{O}, \mathbf{E}\widehat{W}_t\rangle \approx \sum_{k\le K}\int \mathrm{d}\xi\mathrm{d}a\mathrm{d}b\, e^{it\lambda^2 b}\left(\prod_{j=1}^{k+1}\int \frac{-2\pi i F^{(j)}(\xi, v_j)\,\delta(e(v_j)-a)}{b + \lambda^{-2}\varepsilon(\nabla e)(v_j)\cdot\xi - 2i\mathcal{I}(a)}\mathrm{d}v_j\right)$$

with $F^{(1)} := \widehat{W}_0$, $F^{(k+1)} := \mathcal{O}$, and $F^{(j)} \equiv 1$, $j \neq 1, k+1$. Here W_0 is the rescaled Wigner transform of the initial state.

Let $\mathrm{d}\mu_a(v)$ be the normalized surface measure of the level surface $\Sigma_a := \{e(v) = a\}$ defined via the co-area formula, i.e. the integral of any function h w.r.t. $\mathrm{d}\mu_a(v)$ is given by

$$\langle h \rangle_a := \int h(v) \mathrm{d}\mu_a(v) := \frac{\pi}{\mathcal{I}(a)} \int_{\Sigma_a} h(q) \frac{\mathrm{d}m(q)}{|\nabla e(q)|}$$

where $\mathrm{d}m(q)$ is the Lebesgue measure on the surface Σ_a. Often this measure is denoted by $\delta(e(v) - a)\mathrm{d}v$, i.e.

$$\int h(v) \mathrm{d}\mu_a(v) = \frac{\pi}{\mathcal{I}(a)} \int h(v) \delta(e(v) - a) \mathrm{d}v.$$

Let $H(v) := \frac{\nabla e(v)}{2\mathcal{I}(a)}$. Then we have

$$\langle \mathcal{O}, \mathbf{E}\widehat{W}_t \rangle \approx 2\mathcal{I}(a) \sum_{k \le K} \int \mathrm{d}\xi \int_{\mathbb{R}} \mathrm{d}a \mathrm{d}b \, e^{i2t\lambda^2 \mathcal{I}(a) b} \left(\prod_{j=1}^{k+1} \int \frac{-i F^{(j)}(\xi, v_j)}{b + \lambda^{-2} \varepsilon H(v_j) \cdot \xi - i} \mathrm{d}\mu_a(v_j) \right).$$

We expand the denominator up to second order

$$\begin{aligned} &\int \frac{-i}{b + \varepsilon\lambda^{-2} H(v) \cdot \xi - i} \, \mathrm{d}\mu_a(v) \qquad (11.2) \\ &= \frac{-i}{b-i} \int \left[1 - \frac{\varepsilon\lambda^{-2} H(v) \cdot \xi}{b-i} + \frac{\varepsilon^2 \lambda^{-4} [H(v) \cdot \xi]^2}{(b-i)^2} + O\Big((\varepsilon\lambda^{-2} |\xi|)^3 \Big) \right] \mathrm{d}\mu_a(v). \end{aligned}$$

After summation over k, and recalling that $K = O(1)$ due the decay in the observable $\mathcal{O}$, the effect of the last (error) term is $K(\varepsilon\lambda^{-2})^3 = \lambda^{-\kappa}(\lambda^{\kappa/2})^3 = o(1)$, thus we can keep only the first three terms on the right hand side of (11.2).

The linear term cancels out by symmetry: $H(v) = -H(-v)$. To compute the quadratic term, we define

$$D(a) := 4\mathcal{I}(a) \int \mathrm{d}\mu_a(v) \; H(v) \otimes H(v),$$

which is exactly the diffusion matrix (6.7). Thus

$$\begin{aligned} \langle \mathcal{O}, \mathbf{E}\widehat{W}_t \rangle \approx \sum_{k \le K} \int \mathrm{d}\xi \int_{\mathbb{R}} \mathrm{d}a \; 2\mathcal{I}(a) \int \widehat{W}_0(\varepsilon\xi, v_1) \mathrm{d}\mu_a(v_1) \int \mathcal{O}(\xi, v) \mathrm{d}\mu_a(v) \\ \times \int_{\mathbb{R}} \mathrm{d}b \, e^{2i\lambda^2 \mathcal{I}(a) t b} \left(\frac{-i}{b-i} \right)^{k+1} \left[1 + \frac{\varepsilon^2 \lambda^{-4} \langle \xi, D(a)\xi \rangle}{4\mathcal{I}(a)} \frac{1}{(b-i)^2} \right]^{k-1}. \end{aligned}$$

Setting

$$B^2 := \frac{\varepsilon^2 \lambda^{-4} \langle \xi, D(a)\xi \rangle}{4\mathcal{I}(a)},$$

the arising geometric series can be summed up:

$$\sum_{k=0}^{\infty} \left(\frac{-i}{b-i} \right)^{k+1} \left[1 + \frac{B^2}{(b-i)^2} \right]^{k+1} = (-i) \frac{(b-i)^2 + B^2}{(b-i)^3 + i(b-i)^2 + iB^2}.$$

We now perform the db integration analytically by using the formula (with $A := 2\lambda^2\mathcal{I}(a)$):

$$(-i)\int_{\mathbb{R}} \mathrm{d}b\; e^{itAb}\,\frac{(b-i)^2+B^2}{(b-i)^3+i(b-i)^2+iB^2} = 2\pi e^{-tAB^2}+o(1)$$

from the dominant residue $b = iB^2$. We can then compute

$$tAB^2 = \varepsilon^2\lambda^{-4-\kappa}\frac{T}{2}\langle\xi, D(a)\xi\rangle = \frac{T}{2}\langle\xi, D(a)\xi\rangle.$$

In particular, this formula shows that to get a nontrivial limit, ε has to be chosen as $\varepsilon = O(\lambda^{-2-\kappa/2})$, i.e. the diffusive scaling emerges from the calculation. Finally, we have

$$\langle\mathcal{O}, \mathbf{E}\widehat{W}\rangle \approx \int \mathrm{d}\xi \int_{\mathbb{R}} \mathrm{d}a\; \mathcal{I}(a)\Big(\int \mathcal{O}(\xi, v)\mathrm{d}\mu_a(v)\Big)\langle\widehat{W}_0\rangle_a \exp\Big(-\frac{T}{2}\langle\xi, D(a)\xi\rangle\Big)$$

where

$$f(T,\xi,a) := \langle\widehat{W}_0\rangle_a \exp\Big(-\frac{T}{2}\langle\xi, D(a)\xi\rangle\Big)$$

is the solution of the heat equation (6.6) in Fourier space. This completes the sketch of the calculation of the main term and hence the proof of Theorem 8.3. □

12 Conclusions

As a conclusion, we informally summarize the main achievements.

(1) We rigorously proved diffusion from a Hamiltonian quantum dynamics in an environment of fixed time independent random scatterers in the weak coupling regime. The quantum dynamics is described up to a time scale $t \sim \lambda^{-2-\kappa}$ with some $\kappa > 0$, i.e. well beyond the kinetic time scale. The typical number of collisions converges to infinity.

(2) We identified the quantum diffusion constant (or matrix) and we showed that it coincides with the diffusion constant (matrix) obtained from the long time evolution of the Boltzmann equation. This shows that the two-step limit (kinetic limit of quantum dynamics followed by the scaling limit of the random walk) yields the same result, up to the leading order, as the one-step limit (diffusive limit of quantum dynamics).

(3) We controlled the interferences of random waves in a multiple scattering process with infinitely many collisions with random obstacles in the extended states regime.

(4) In agreement with (2), we showed that quantum interferences and memory effects do not become relevant up to our scale. We remark that this is expected to hold for any κ in $d = 3$, but *not expected* to hold for $d = 2$ [localization].

(5) As our main technical achievement, we classified and estimated Feynman graphs up to *all orders*. We gained an extra λ-power *per each non-ladder vertex* compared with the usual power counting estimate relying on the L^∞ and L^1-bounds for the propagators and on the tree and loop momenta.

A Inequalities

In this appendix we collect the precise bounds on integrals of propagators that are used in the actual proofs. Their proofs are more or less elementary calculations, however the calculations are much easier for the continuous dispersion relation. For details, see the Appendix B of [24] for the continuous model and in Appendix A of [25] for the discrete case.

We define the weighted Sobolev norm

$$\|f\|_{m,n} := \sum_{|\alpha|\le n} \|\langle x\rangle^m \partial^\alpha f(x)\|_\infty$$

with $\langle x\rangle := (2+x^2)^{1/2}$ (here α is a multiindex).

Lemma A.1 (Continuous dispersion relation [24]). *Let $e(p) = \frac{1}{2}p^2$. Suppose that $\lambda^2 \ge \eta \ge \lambda^{2+4\kappa}$ with $\kappa \le 1/12$. Then we have,*

$$\int \frac{|h(p-q)|\mathrm{d}p}{|\alpha-\omega(p)+i\eta|} \le \frac{C\|h\|_{2d,0}\,|\log\lambda|\,\log\langle\alpha\rangle}{\langle\alpha\rangle^{1/2}\langle |q|-\sqrt{2|\alpha|}\rangle}\,, \tag{A.1}$$

and for $0 \le a < 1$

$$\int \frac{|h(p-q)|\mathrm{d}p}{|\alpha-e(p)+i\eta|^{2-a}} \le \frac{C_a\|h\|_{2d,0}\,\eta^{-2(1-a)}}{\langle\alpha\rangle^{a/2}\langle |q|-\sqrt{2|\alpha|}\rangle}\,. \tag{A.2}$$

For $a=0$ and with $h := \widehat{B}$, the following more precise estimate holds. There exists a constant C_0, depending only on finitely many $\|B\|_{k,\kappa}$ norms, such that

$$\int \frac{\lambda^2|\widehat{B}(p-q)|^2\,\mathrm{d}p}{|\alpha-\overline{\omega}(p)-i\eta|^2} \le 1 + C_0\lambda^{-12\kappa}\big[\lambda + |\alpha-\omega(q)|^{1/2}\big]\,. \tag{A.3}$$

One can notice that an additional decaying factor $h(p-q)$ must be present in the estimates due to the possible (but irrelevant) ultraviolet divergence. In the applications $h = \widehat{B}$. Moreover, an additional decaying factor in α was also saved, this will help to eliminate the logarithmic divergence of the integral of the type

$$\int_{\mathbb{R}} \frac{\mathrm{d}\alpha}{|\alpha-c+i\eta|} \sim |\log\eta|$$

modulo the logarithmic divergence at infinity.

The following statement estimates singularity overlaps:

Lemma A.2. *Let $\mathrm{d}\mu(p) = \mathbf{1}(|p|\le\zeta)\mathrm{d}p$, in applications $\zeta \sim \lambda^{-\kappa}$. For any $|q| \le C\lambda^{-1}$*

$$I_1 := \int \frac{\mathrm{d}\mu(p)}{|\alpha-e(p)+i\eta|\,|\beta-e(p+q)+i\eta|} \le \frac{C\zeta^{d-3}|\log\eta|^2}{\|\!|q|\!\|} \tag{A.4}$$

$$I_2 := \int \frac{\mathrm{d}\mu(p)}{|\alpha-e(p)+i\eta|\,|\beta-e(p+q)+i\eta|}\,\frac{1}{\|\!|p-r|\!\|} \le \frac{C\eta^{-1/2}\zeta^{d-3}|\log\eta|^2}{\|\!|q|\!\|} \tag{A.5}$$

uniformly in r,α,β. Here $\|\!|q|\!\| := |q| + \eta$.

To formulate the statement in the discrete case, we redefine

$$\|\!|q|\!\| := \eta + \min\{|q-\gamma| \ : \ \gamma \text{ is a critical point of } e(p)\}$$

for any momentum q on the d-dimensional torus. It is easy to see that there are 2^d critical points of the discrete dispersion relation $e(p) = \sum_{j=1}^d (1-\cos p^{(j)})$.

Lemma A.3 (Discrete dispersion relation [25])**.** *The following bounds hold for the dispersion relation $e(p) = \sum_{j=1}^d (1-\cos p^{(j)})$ in $d \geq 3$ dimensions.*

$$\int \frac{\mathrm{d}p}{|\alpha - e(p) + i\eta|} \frac{1}{\|\!|p-r|\!\|} \leq c|\log\eta|^3 \,, \tag{A.6}$$

$$I := \int \frac{\mathrm{d}p}{|\alpha - e(p) + i\eta|} \frac{1}{|\beta - e(p+q) + i\eta|} \leq \frac{c\eta^{-3/4}|\log\eta|^3}{\|\!|q|\!\|} \tag{A.7}$$

$$I(r) := \int \frac{\mathrm{d}p}{|\alpha - e(p) + i\eta|} \frac{1}{|\beta - e(p+q) + i\eta|} \frac{1}{\|\!|p-r|\!\|} \leq \frac{c\eta^{-7/8}|\log\eta|^3}{\|\!|q|\!\|} \tag{A.8}$$

uniformly in r, α, β. With the (carefully) renormalized dispersion relation it also holds that

$$\sup_\alpha \int \frac{\lambda^2 \,\mathrm{d}p}{|\alpha - \overline{\omega}(p) - i\eta|^2} \leq \frac{1}{\eta} + C_0\lambda^{1-12\kappa} \tag{A.9}$$

(compare with (A.3)*).*

Because the gain in the integrals (A.7), (A.8) compared with the trivial estimate $\eta^{-1}|\log\eta|$, is quite weak (compare with the much stronger bounds in Lemma A.2 for the continuous case), we need one more inequality that contains four denominators. This special combination occurs in the estimates of the recollision terms. The proof of this inequality is considerably more involved than the above ones, see [19]. The complication is due to the lack of convexity of the level sets of the dispersion relation (see Fig. 21). We restrict our attention to the $d = 3$ dimensional case, as it turns out, this complication is less serious in higher dimensions.

For any real number α we define

$$\|\!|\alpha|\!\| := \min\{|\alpha|, |\alpha-2|, |\alpha-3|, |\alpha-4|, |\alpha-6|\} \tag{A.10}$$

in the $d = 3$ dimensional model. The values $0, 2, 4, 6$ are the critical values of $e(p)$. The value $\alpha = 3$ is special, for which the level surface $\{e(p) = 3\}$ has a flat point. In general, in $d \geq 3$ dimensions, $\|\!|\alpha|\!\|$ is the minimum of $|\alpha - d|$ and of all $|\alpha - 2m|$, $0 \leq m \leq d$.

Lemma A.4. *[Four Denominator Lemma [19]] For any $\Lambda > \eta$ there exists C_Λ such that for any $\alpha \in [0,6]$ with $\|\!|\alpha|\!\| \geq \Lambda$,*

$$\mathcal{I} = \int \frac{\mathrm{d}p\mathrm{d}q\mathrm{d}r}{|\alpha - e(p) + i\eta||\alpha - e(q) + i\eta||\alpha - e(r) + i\eta||\alpha - e(p-q+r+u) + i\eta|} \leq C_\Lambda|\log\eta|^{14} \tag{A.11}$$

uniformly in u.

The key idea behind the proof of Lemma A.4 is to study the decay property of the Fourier transform of the level sets $\Sigma_a := \{p \ : \ e(p) = a\}$, i.e. the quantity

$$\widehat{\mu}_a(\xi) = \int_{\Sigma_a} e^{ip\xi} \mathrm{d}p := \int_{\Sigma_a} \frac{e^{ip\xi}}{|\nabla e(p)|} \mathrm{d}m_a(p),$$

where m_a is the uniform (euclidean) surface measure on Σ_a. Defining

$$I(\xi) = \int \frac{e^{ip\xi}\mathrm{d}p}{|\alpha - e(p) + i\eta|},$$

we get by the coarea formula that

$$I(\xi) = \int_0^6 \frac{\mathrm{d}a}{|\alpha - a + i\eta|} \widehat{\mu}_a(\xi).$$

From the convolution structure of $\mathcal{I}$, we have

$$\mathcal{I} = \int I(\xi)^4 e^{-iu\xi} \mathrm{d}\xi \leq \int |I(\xi)|^4 \mathrm{d}\xi \leq \Big(\int_0^6 \frac{\mathrm{d}a}{|\alpha - a + i\eta|} \Big)^4 \sup_a \int |\widehat{\mu}_a(\xi)|^4 \mathrm{d}\xi.$$

The first factor on the right hand side is of order $|\log \eta|^4$, so the key question is the decay of the Fourier transform, $\widehat{\mu}_a(\xi)$, of the level set, for large ξ. It is well known (and follows easily from stationary phase calculation) that for surfaces $\Sigma \subset \mathbb{R}^3$ whose Gauss curvature is strictly separated away from zero (e.g. strictly convex surfaces), the Fourier transform

$$\widehat{\mu}_\Sigma(\xi) := \int_\Sigma e^{ip\xi} \, \mathrm{d}p$$

decays as

$$|\widehat{\mu}_\Sigma(\xi)| \leq \frac{C}{|\xi|}$$

for large ξ. Thus for such surfaces the L^4-norm of the Fourier transform is finite. For surfaces, where only one principal curvature is non-vanishing, the decay is weaker, $|\xi|^{-1/2}$. Our surface Σ_a even has a flat point for $a = 3$. A detailed analysis shows that although the decay is not sufficient uniformly in all direction ξ, the exceptional directions and their neighborhoods are relatively small so that the L^4-norm is still finite. The precise statement is more involved, here we flash up the main result. Let $K = K(p)$ denote the Gauss curvature of the level set Σ_a and let $\nu(p)$ denote the outward normal direction at $p \in \Sigma_a$.

Lemma A.5 (Decay of the Fourier transform of Σ_a, [19]). *For $\nu \in S^2$ and $r > 0$ we have*

$$\widehat{\mu}_a(r\nu) \lesssim \frac{1}{r} + \frac{1}{r^{3/4}|D(\nu)|^{1/2} + 1},$$

where $D(\nu) = \min_j |\nu(p_j) - \nu|$ where p_j's are finitely many points on the curve $\Gamma = \{K = 0\} \cap \Sigma_a$, at which the neutral direction of the Gauss map is parallel with Γ.

This lemma provides sufficient decay for the L^4-norm to be only logarithmically divergent, which can be combined with a truncation argument to obtain (A.11).

We stated all above estimates (with the exception of the precise (A.3) and (A.9)) for the bare dispersion relation $e(p)$. The modification of these estimates for the renormalized dispersion relation $\omega(p)$ is straightforward if we are prepared to lose an additional $\eta^{-\kappa}$ factor by using

$$\frac{1}{|\alpha-\omega(p)+i\eta|} \le \frac{1}{|\alpha-e(p)+i\eta|} + \frac{c\lambda^2}{|\alpha-e(p)+i\eta||\alpha-\omega(p)+i\eta|} \le \frac{1+c\lambda^2\eta^{-1}}{|\alpha-e(p)+i\eta|} .$$

B Power counting for integration of Feynman diagrams

Lemma B.1. *Let Γ be a connected oriented graph with set of vertices $V(\Gamma)$ and set of edges $E(\Gamma)$. Let $N = |V(\Gamma)|$ be the number of vertices and $K = |E(\Gamma)|$ be the number of edges. Let $p_e \in \mathbb{R}^d$ denote a (momentum) variable associated with the edge $e \in E(\Gamma)$ and $\mathbf{p}$ denotes the collection $\{p_e \,:\, e \in E(\Gamma)\}$. Since the graph is connected, $K \ge N-1$. Let $R : \mathbb{R}^d \to \mathbb{C}$ denote a function with finite L^1 and L^∞ norms, $\|R\|_1$ and $\|R\|_\infty$. For any vertex $w \in V(\Gamma)$, let*

$$\Omega_w := \int \Delta_w(\mathbf{p}) \prod_{e\in E(\Gamma)} R(p_e)\mathrm{d}p_e,$$

where

$$\Delta_w(\mathbf{p}) = \prod_{\substack{v\in V(\Gamma)\\ v\neq w}} \delta\Big(\sum_{e\,:\,e\sim v} \pm p_e\Big) \tag{B.12}$$

is a product of delta functions expressing Kirchoff laws at all vertices but w. Here $e \sim v$ denotes the fact that the edge e is adjacent to the vertex v and $\pm$ indicates that the momenta have to be summed up according to the orientation of the edges with respect to the vertex. More precisely, if the edge e is outgoing with respect to the vertex v, then the sign is minus, otherwise it is plus.

Then the integral Ω_w is independent of the choice of the special vertex w and the following bound holds

$$|\Omega_w| \le \|R\|_1^{K-N+1}\|R\|_\infty^{N-1}.$$

Proof. First notice that the arguments of the delta functions for *all* $v \in V(\Gamma)$,

$$\sum_{e\,:\,e\sim v} \pm p_e,$$

sum up to zero (every p_e, $e \in E(\Gamma)$, appears exactly twice, once with plus once with minus). This is the reason why one delta function had to be excluded from the product (B.12), otherwise they would not be independent. It is trivial linear algebra to see that, independently of the choice of w, all Δ_w determine the same linear subspace in the space of momenta, $(\mathbb{R}^d)^{E(\Gamma)} = \mathbb{R}^{dK}$.

Let T be a spanning tree in Γ and let $\mathcal{T} = E(T)$ denote the edges in Γ that belong to T. Let $\mathcal{L} = E(\Gamma) \setminus \mathcal{T}$ denote the remaning edges, called "loops". Momenta associated with T,

$\{p_e \ : \ e \in \mathcal{T}\}$, are called tree-momenta, the rest are called loop-momenta. Under Kirchoff law at the vertices, all tree momenta can be expressed in terms of linear combinations (with coefficients ± 1) of the loop momenta:

$$p_e = \sum_{a \in \mathcal{L}} \sigma_{e,a} p_a, \qquad \forall e \in \mathcal{T}, \tag{B.13}$$

where the coefficients $\sigma_{e,a}$ can be determined as follows. Consider the graph $T \cup \{a\}$, i.e. adjoin the edge a to the spanning tree. This creates a single loop L_a in the graph that involves some elements of $\mathcal{T}$. We set $\sigma_{e,a} = 1$ if $e \in L_a$ and the orientation of e and a within this loop coincide. We set $\sigma_{e,a} = -1$ if the orientation of e and a are opposite, and finally $\sigma_{e,a} = 0$ if $e \notin L_a$.

It is easy to check that the $N-1$ linear relations (B.13) are equivalent to the ones determined by the product (B.12) of delta functions. For, if all tree momenta are defined as linear combinations of the loop momenta given by (B.13), then the Kirchoff law is trivially satisfied. To check this, notice that when summing up the edge momenta at each vertex, only those loop momenta p_a appear whose loop L_a contains v. If $v \in L_a$, then p_a appears exactly twice in the sum, once with plus and once with minus, since the momenta p_a flowing along the loop L_a once enters and once exits at the vertex v. Thus the relations (B.13) imply the relations in (B.12), and both sets of relations have the same cardinality, $N-1$. Since the relations (B.13) are obviously independent (the tree momenta are all different), we obtain that both sets of relations determine the same subspace (of codimension $d(N-1)$) of $\mathbb{R}^{dK}$.

Thus we can rewrite

$$\Omega = \Omega_w = \int \left(\prod_{e \in E(\Gamma)} R(p_e) \mathrm{d}p_e \right) \prod_{e \in \mathcal{T}} \delta\big(p_e - \sum_{a \in \mathcal{L}} \sigma_{e,a} p_a\big);$$

in particular, Ω_w is independent of w.

Now we estimate all factors $R(p_e)$ with $e \in \mathcal{T}$ by L^∞-norm, then we integrate out all tree momenta and we are left with the L^1-norms of the loop momenta:

$$|\Omega| \le \|R\|_\infty^{N-1} \int \Big(\prod_{e \in \mathcal{L}} |R(p_e)| \mathrm{d}p_e \Big) \Big(\prod_{e \in \mathcal{T}} \delta\big(p_e - \sum_{a \in \mathcal{L}} \sigma_{e,a} p_a\big) \mathrm{d}p_e \Big) = \|R\|_\infty^{N-1} \|R\|_1^{K-N+1}$$

completing the proof of Lemma B.1. □

References

[1] R. Adami, L. Erdős, *Rate of decoherence for an electron weakly coupled to a phonon gas.* J. Statis. Physics **132**, no. 2, 301–328 (2008)

[2] M. Aizenman and S. Molchanov, *Localization at large disorder and at extreme energies: an elementary derivation*, Commun. Math. Phys. **157**, 245–278 (1993)

[3] M. Aizenman, R. Sims, S. Warzel, *Absolutely continuous spectra of quantum tree graphs with weak disorder.* Comm. Math. Phys. **264**, 371-389 (2006)

[4] P. Anderson, *Absences of diffusion in certain random lattices*, Phys. Rev. **109**, 1492–1505 (1958)

[5] D. Benedetto, F. Castella, R. Esposito and M. Pulvirenti, *From the N-body Schrodinger equation to the quantum Boltzmann equation: a term-by-term convergence result in the weak coupling regime.* Commun. Math. Phys. **277**, 144 (2008)

[6] R. Brown, Philosophical Magazine N. S. **4** 161–173 (1828), and **6** 161–166 (1829)

[7] J. Bourgain, *Random lattice Schrödinger operators with decaying potential: some higher dimensional phenomena.* Lecture Notes in Mathematics, Vol. 1807, 70-99 (2003).

[8] L. Bunimovich, Y. Sinai: *Statistical properties of Lorentz gas with periodic configuration of scatterers.* Commun. Math. Phys. **78** no. 4, 479–497 (1980/81),

[9] T. Chen, *Localization Lengths and Boltzmann Limit for the Anderson Model at Small Disorders in Dimension 3.* J. Stat. Phys. **120**, 279–337 (2005)

[10] W. De Roeck, J. Fröhlich: *Diffusion of a massive quantum particle coupled to a quasi-free thermal medium in dimension $d \geq 4$.* Preprint. http://arxiv.org/abs/0906.5178

[11] W. De Roeck, J. Fröhlich, A. Pizzo: *Quantum Brownian motion in a simple model system.* Comm. Math. Phys. **293** (2010), no. 2, 361–398.

[12] Disertori, M., Spencer, T.: *Anderson localization for a supersymmetric sigma model.* Preprint arXiv:0910.3325.

[13] Disertori, M., Spencer, T., Zirnbauer, M.: *Quasi-diffusion in a 3D Supersymmetric Hyperbolic Sigma Model.* Preprint arXiv:0901.1652.

[14] D. Dürr, S. Goldstein, J. Lebowitz: *A mechanical model of Brownian motion.* Commun. Math. Phys. **78** (1980/81) no. 4, 507-530.

[15] A. Einstein, *Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen*, Ann. der Physik, **17**, 549–560 (1905), and: *Zur Theorie der Brownschen Bewegung*, Ann. der Physik, **19** 180 (1906).

[16] K. Efetov. Supersymmetry in Disorder and Chaos, Cambridge University Press, 1999.

[17] D. Eng and L. Erdős, *The Linear Boltzmann Equation as the Low Density Limit of a Random Schrödinger Equation.* Rev. Math. Phys, **17**, No. 6, 669-743 (2005)

[18] L. Erdős, *Linear Boltzmann equation as the long time dynamics of an electron weakly coupled to a phonon field.* J. Stat. Phys., **107**, 1043-1128 (2002)

[19] L. Erdős, M. Salmhofer, *Decay of the Fourier transform of surfaces with vanishing curvature.* Math. Z. **257** no 2., 261-294 (2007)

[20] L. Erdős and H.-T. Yau, *Linear Boltzmann equation as scaling limit of quantum Lorentz gas.* Advances in Differential Equations and Mathematical Physics. Contemporary Mathematics **217**, 137-155 (1998).

[21] L. Erdős and H.-T. Yau, *Linear Boltzmann equation as the weak coupling limit of the random Schrödinger equation*, Commun. Pure Appl. Math. **LIII**, 667-735, (2000).

[22] L. Erdős, M. Salmhofer and H.-T. Yau, *Towards the quantum Brownian motion.* Lecture Notes in Physics, **690**, Mathematical Physics of Quantum Mechanics, Selected and Refereed Lectures from QMath9. Eds. Joachim Asch and Alain Joye. pp. 233-258 (2006)

[23] L. Erdős, M. Salmhofer, H.-T. Yau, *Quantum diffusion of the random Schrodinger evolution in the scaling limit.* Acta Math. **200**, no.2, 211-277 (2008)

[24] L. Erdős, M. Salmhofer, H.-T. Yau, *Quantum diffusion of the random Schrodinger evolution in the scaling limit II. The recollision diagrams.* Commun. Math. Phys. **271**, 1-53 (2007)

[25] L. Erdős, M. Salmhofer, H.-T. Yau, *Quantum diffusion for the Anderson model in scaling limit.* Ann. Inst. H. Poincare **8**, 621-685 (2007)

[26] R. Froese, D. Hasler, W. Spitzer, *Absolutely continuous spectrum for the Anderson model on a tree: a geometric proof of Klein's theorem.* Comm. Math. Phys. **269**, 239-257 (2007)

[27] J. Fröhlich, F. Martinelli, E. Scoppola and T. Spencer, *Constructive proof of localization in the Anderson tight binding model.* Comm. Math. Phys. **101**, no. 1, 21-46 (1985)

[28] J. Fröhlich and T. Spencer, *Absence of diffusion in the Anderson tight binding model for large disorder or low energy*, Commun. Math. Phys. **88**, 151–184 (1983)

[29] F. Germinet, A. Klein, J. Schenker: *Dynamical delocalization in random Landau Hamiltonians.* Ann. of Math. **166**, 215-244 (2007)

[30] T. G. Ho, L. J. Landau and A. J. Wilkins: *On the weak coupling limit for a Fermi gas in a random potential.* Rev. Math. Phys. **5** (1993), no. 2, 209-298.

[31] I. Ya. Goldsheid, S. A. Molchanov and L. A. Pastur, *A pure point spectrum of the one dimensional Schrödinger operator.* Funct. Anal. Appl. **11**, 1–10 (1997)

[32] Y. Kang, J. Schenker: *Diffusion of wave packets in a Markov random potential.* J. Stat. Phys. **134** (2009), 1005-1022.

[33] H. Kesten, G. Papanicolaou: *A limit theorem for stochastic acceleration.* Comm. Math. Phys. **78** 19-63. (1980/81)

[34] A. Klein, *Absolutely continuous spectrum in the Anderson model on the Bethe lattice*, Math. Res. Lett. **1**, 399–407 (1994)

[35] T. Komorowski, L. Ryzhik: *Diffusion in a weakly random Hamiltonian flow.* Comm. Math. Phys. **263**, 277-323 (2006)

[36] O. Lanford, *On the derivation of the Boltzmann equation.* Astérisque **40**, 117-137 (1976)

[37] P. A. Lee, T. V. Ramakrishnan, *Disordered electronic systems.* Rev. Mod. Phys. **57**, 287–337 (1985)

[38] J. Lukkarinen and H. Spohn, *Kinetic limit for wave propagation in a random medium.* Arch. Ration. Mech. Anal. **183**, 93-162 (2007)

[39] J. Lukkarinen and H. Spohn, *Weakly nonlinear Schrödinger equation with random initial data.* Preprint arXiv:0901.3283

[40] P. Markowich and C. Ringhofer, *Analysis of the quantum Liouville equation.* Z. Angew. Math. Mech, **69**, 121-129 (1989).

[41] J. B. Perrin, *Mouvement brownien et réalité moléculaire.* Annales de chimie et de physiqe, VIII 18, 5–114 (1909).

[42] C.–A. Pillet, *Some results on the quantum dynamics of a particle in a Markovian potential.* Comm. Math. Phys. **102** (1985), no. 2, 237254.

[43] B. Simon and T. Wolff, *Singular continuous spectrum under rank one perturbations and localization for random Hamiltonians.* Comm. Pure Appl. Math. **39**(1), 75-90 (1986)

[44] M. von Smoluchowski, *Zur kinetischen Theorie der Bwownschen Molekularbewegung und der Suspensionen.* Ann. der Physik, **21**, 756–780 (1906).

[45] H. Spohn: *Derivation of the transport equation for electrons moving through random impurities.* J. Statist. Phys. **17** (1977), no. 6., 385-412.

[46] D. Vollhardt, P. Wölfle, *Diagrammatic, self-consistent treatment of the Anderson localization problem in $d \leq 2$ dimensions.* Phys. Rev. B **22**, 4666-4679 (1980)

[47] N. Wiener, *Differential space.* J. Math and Phys. **58**, 131-174 (1923)