

1

Basic issues in nonequilibrium statistical mechanics

Perhaps due to its technical complexity, oftentimes one sees in research papers on nonequilibrium quantum field theory (NEqQFT) more emphasis placed on the field-theoretical formalisms than the ideas these sophisticated techniques attempt to capture, or the issues such problems embody. All the more so, we need some basic understanding of the important issues and concepts in nonequilibrium statistical mechanics (NEqSM), and how they are manifested in the context of quantum field theory. Many important advances in this field came from asking such questions and finding out how to answer them in the language of quantum field theory. Because of this somewhat skewed existing emphasis in NEqQFT, and since we do not assume the reader to have had a formal course on NEqSM before, we shall give a brief summary of the basic concepts of NEqSM relevant to the field-theoretical processes discussed in this book. Many fine monographs and reviews written on this subject take a more formal mathematical approach. Since our purpose here is to familiarize readers with these issues and their subtleties, rather than training them to work in the rich field of NEqSM (which includes in addition to the traditional subject matter such as the projection operator formalism and open system concepts, also current topics at their foundation, such as dynamical systems and quantum chaos), we choose to approach these topics in a more intuitive and physical way, sacrificing by necessity rigor and completeness.

We first examine some commonly encountered physical processes and try to bring out in each a different key concept in NEqSM. To have a concrete bearing and a common ground, let us focus on just one such issue which is of paramount importance and poses a constant challenge to theoretical physicists: How does apparent irreversibility in the macroscopic world arise from the time-reversal invariant laws of microphysics [Leb93, HaPeZu94, Mac92, Sch97]?

We begin with an analysis of the nature and origin of irreversibility in well-known physical processes such as dispersion (referring in the specific context here to the divergence of neighboring trajectories in configuration or phase space due to dynamical instability), diffusion, dissipation and mixing. We will seek the microdynamical basis of these processes and clarify the distinction between processes whose irreversibility arises from the stipulation of special initial conditions, and those arising from the system's interaction with a coarse-grained environment. It is beneficial to keep in mind these processes and the issues they embody when we begin our study of quantum field processes so that they will not be marred by the technical complexity of quantum field theory. We

can ask questions such as (1) “What is the entropy generation from particle creation in an external field or a dynamical spacetime, as in cosmology?”; (2) “How could an interacting field thermalize?”; (3) “Is there irreversibility associated with quantum fluctuations in field theoretical processes like particle creation?” Or, more boldly, “Can the ‘birth of the Universe’ be viewed as a large fluctuation?” “Might it not happen at all – a ‘still’ birth – due to the powerful dissipative effects of particle creation which suppresses the tunneling rate?” (4) “Can one use thermodynamic relations to characterize certain quantum field processes?”

These questions reveal how deeply one can probe into the NEqSM features of quantum field theory and how quantum field processes can lend themselves to statistical mechanical and thermodynamic depiction or characterization. Asking question (1) reveals the differences resulting from many levels of coarse graining between a quantum field understanding of particle creation processes (no entropy production because the vacuum is a pure state) and a thermodynamic description (yes, entropy is proportional to the number of particles produced). Asking question (2) forces us to reckon with the intricate NEqSM features of an interacting quantum field such as how a correlation entropy can be defined from the Schwinger–Dyson hierarchy. These aspects are not usually discussed in quantum field theory textbooks. The first part of question (3) brings out the often used yet poorly understood aspects of noise – beginning with quantum noise associated with vacuum fluctuations, properties of multiplicative colored noise, and nonlocal dissipation and their effects on the dynamical processes. The second part of question (3) is the so-called “back-reaction” effect of quantum fields on a background field or background spacetime. Question (4) asks if this effect can have a thermodynamic interpretation. To the degree that thermodynamics is the long-wavelength, heavily coarse-grained limit of microphysics and quantum field theory is a theory of microphysics, we certainly expect such relations to exist and their discovery will reveal the relation between micro–macro and quantum-to-classical transitions. A well-known relation is the black hole thermodynamics of Bekenstein [Bek73] and the quantum Hawking radiation [Haw75]. Sciama [Sci79] suggested that this can be understood from the viewpoint of quantum dissipative systems. This view also applies to dissipation of anisotropy in the early universe due to particle creation from the vacuum. We will find out later that both for the black hole and the early universe these processes can indeed be understood as manifestations of a fluctuation–dissipation relation, relating fluctuations of quantum fields to dissipation in the dynamics of the background field or spacetime.

1.1 Macroscopic description of physical processes

Let us begin by examining a few examples of irreversible processes to illustrate their different natures and origins. Consider the following processes:

Dispersion
Diffusion
Dissipation
Relaxation
Mixing
Recurrence
Decoherence
Recoherence

They contain different aspects of irreversibility. The usage of these terms appearing in general-purpose books could be rather loose or even confusing. For example, diffusion, relaxation and dissipation are often seen used interchangeably. Even the same word could mean different things in different contexts. For example, classical diffusion is often viewed as a form of dissipation, while quantum diffusion refers to phase dispersion, usually occurs at a much faster time-scale and is more closely related to decoherence than dissipation. We will discuss quantum phenomena in Chapter 3. Here we will focus on the first six such processes listed above and aim at providing some microdynamics basis to these processes in order to give them a more precise meaning. In so doing we hope to elucidate some basic notions and issues of NEqSM through examples.

We first highlight the distinction between dissipative processes (which are always irreversible) and irreversible or “apparently” irreversible processes (which are not necessarily dissipative). For example, in elastic scattering, neighboring trajectories diverge exponentially fast. This is characteristic of mixing systems, which are reversible. Relaxation and diffusion referring to dissipative systems are irreversible. They are mixing systems with some type of coarse graining introduced. As we shall see, not any type of coarse graining leads to irreversibility. Many factors enter, such as the large size of the system, the particular initial conditions chosen, or the time-scales at work. This is where it calls for special caution in doing the analysis. Better understanding of the chaotic behavior in classical molecular dynamics has provided a firmer microscopic basis for nonequilibrium statistical mechanics. Such studies for quantum systems are less developed and for this reason we shall refrain from describing them. In Chapter 3 we shall have occasion to discuss quantum decoherence and dissipation where the interplay of quantum and thermal fluctuations in the environment and their effects on the system will be discussed. We shall also revisit these issues of irreversibility and approach to equilibrium in Chapter 12.

A. Dispersion

Consider a system of dilute gas made up of interacting particles modeled as hard spheres with diameter d . For simplicity, let us work in two dimensions with hard disks. (Our illustration here follows [Gas98]; see also [Ma85] which contains excellent conceptual discussions.) Assume the particles move with constant velocity v

and traverse a distance given by the mean free path $\ell \gg d$ before colliding with another particle elastically. The trajectory of any particle governed by the laws of mechanics is of course reversible in time. However, upon just a few collisions two neighboring trajectories will deviate from each other very rapidly if the scattering surface is convex, as a sphere is. To see this, let's set our stop watch time zero ($t = 0$) right after the first collision (call this collision the $n = 0$ one) and follow the particle's trajectory for n subsequent collisions. Call the scattering angle of the first collision $\theta(0)$ and the uncertainty associated with it $\delta\theta(0)$ and likewise for the scattering angle after an additional n collisions $\theta(t)$ and its uncertainty $\delta\theta(t)$. For each additional collision the uncertainty in the scattering angle increases by a factor of ℓ/d deduced from the simple trigonometry of incident and scattered trajectories. So after n collisions then

$$|\delta\theta(t)| \sim |\delta\theta(0)|^n \equiv |\delta\theta(0)|e^{\lambda t} \quad (1.1)$$

The second equivalence relation above defines the parameter λ , which is called the Lyapunov exponent (actually its maximal value enters into this expression). The time for n successive collisions is given by $t = n\tau$ where τ is the time between collisions related to the mean free path ℓ by $v = \ell/\tau$. Thus the (maximal) Lyapunov exponent is given by

$$\lambda \sim \frac{1}{\tau} \ln \frac{\ell}{d} \quad (1.2)$$

This simple way of estimating the maximum Lyapunov exponent first given by Krylov [Kry44, Kry79] remains very useful in illustrating the elemental process of divergence of neighboring trajectories due to dynamical instability, referred to here as "dispersion" for short. For hard sphere collisions we see that after a sufficiently long time $|\delta\theta(t)| \approx 1$, the exit direction becomes completely indeterminate due to the accumulated error.

The asymmetry in the initial and final conditions of the collection of trajectories (congruence) comes from the accumulation and magnification of the *uncertainty in the initial conditions* due to the collisions, even though the dynamical law governing each trajectory is time-symmetric. To trace a particular trajectory backwards in time after a large number of collisions requires an exponentially high degree of precision in the specification of the initial condition. This ultra-sensitivity of dynamics to initial conditions is characteristic of chaotic systems. Note that the divergence of neighboring trajectories in phase space or parameter space is an intrinsic property of the nonlinear Hamiltonian of the system, not a result of coarse graining by the truncation of the Bogoliubov–Born–Green–Kirkwood–Yvon (BBGKY) series *and* the causal factorizability of the two-particle correlation function as in Boltzmann's molecular chaos hypothesis. (Initially uncorrelated particles become correlated after collisions, thus giving rise to time-asymmetry in the dissipative dynamics of Boltzmann's equation.) The evolution of an ensemble

of such systems at some finite time from the initial moment often appears to be unrelated to their initial conditions, not because the individual systems are insensitive to the initial conditions but because they are overly-sensitive to them, thus making it difficult to provide an accurate prediction of each system's state in the future. It is in this sense that these systems manifest irreversibility. In contrast, for an integrable system the trajectories stay close to each other because the regions in phase space for its dynamics are limited by the constants of motion. Such trajectories in integrable systems are referred to as “stable” while those in chaotic systems are “unstable” as they become dispersive in the sense defined above owing to their dynamical instability. We will return in a later section to irreversibility and nonequilibrium thermodynamics considered from the framework of Hamiltonian dynamics.

B. Diffusion

Let us look at some simple examples in kinetic theory: gas expansion, ice melting and an ink drop in water. These are irreversible processes because the initial states of 10^{23} molecules on one side of the chamber and a piece of ice or ink drop immersed in a bath of water are *highly improbable configurations* out of all possible arrangements. These initial conditions are states of very low entropy. The only reason why they are special is because we arrange them to be so. For these problems, we also know that the system–environment separation and interaction make a difference in the outcome. In the case of an expanding ideal gas, for example, for free expansion the change of entropy is $\delta S_{\text{system}} > 0$, $\delta S_{\text{environ}} = 0$, $\delta S_{\text{total}} > 0$. For isothermal *quasistatic* expansion: $\delta S_{\text{system}} = -\delta S_{\text{environ}} > 0$, $\delta S_{\text{total}} = 0$ instead (see, e.g. [Rei65]).

Another important factor in determining whether a process is irreversible is the *time-scale* of observation compared to the dynamic time-scale of the process. We are familiar with the irreversible process of an ink drop dispersing in water which happens in a matter of seconds, but the same dye suspension put in glycerine takes days to diffuse, and for a short duration after the initial mixing (say, by cranking the column of glycerine with a vertical stripe of dye one way) one can easily “unmix” them (by reversing the direction of cranking [UMDdemo]). We will discuss in the next section under what conditions and in what sense a “mixing” system, though time-reversible, can be viewed as capable of approaching equilibrium. Diffusion, when used in the sense of dissipation, is nevertheless an irreversible process.

C. Dissipation

There are two basic models of dissipation in nonequilibrium statistical mechanics: the Boltzmann kinetic theory of dilute gas, and the Langevin theory of Brownian motion. Each invokes a different set of concepts, and even their relation is illustrative. In kinetic theory, the equations governing the n -particle distribution functions (the BBGKY hierarchy) preserve the full information of an

n -particle system. It is (1) ignoring (more often restricted by the precision of one's observation than by choice) the information contained in the higher-order correlations (truncation of the BBGKY hierarchy), and (2) the imposition of causal factorization conditions, like the molecular chaos assumption, that brings about dissipation and irreversibility in the dynamics of the lower-order correlations [Zwa01, Bal75].

In the lowest order truncation of the BBGKY hierarchy valid for the description of dilute gases, the Liouvillian operator L acting on the one-particle distribution function $f_1(r_1, p_1, t)$ is driven by a collision integral involving a two-particle distribution function $f_2(r_1, p_1, r_2, p_2, t)$ (cf. Chapters 2 and 11). Boltzmann's molecular chaos ansatz (MCA) assumes an initial uncorrelated state between two particles: $f_2(1, 2) = f_1(1)f_1(2)$, i.e. that the probability of finding particle 1 at (r_1, p_1, t) and particle 2 at (r_2, p_2, t) at the same time t is equal to the product of the single-particle probabilities (a factorizable condition). Note that this condition is assumed to hold only initially, but not finally. A short-range interaction in a collision process will almost certainly generate dynamical correlations between the two collision partners. The truncated BBGKY hierarchy (with MCA) is an example of what we call an effectively open system (see Section 1.5 of this chapter). Boltzmann's explanation of dissipation in macroscopic dynamics is one of the crowning achievements of theoretical physics.

Dissipation in an open system described by the Langevin dynamics has similarities with and differences from that of an effectively open system (as exemplified by the Boltzmann system). The open system can be one distinguished oscillator, the Brownian particle (with mass M), interacting with many oscillators (with mass m) serving as its environment (see Chapter 2). Dissipation in the dynamics of the open system arises from ignoring details of the environmental variables and only keeping their averaged effect on the system (this also brings about a renormalization of the mass and the natural frequency of the Brownian particle). Usually one assumes $M \gg m$ and weak coupling between the system and the environment to simplify calculations. The effect of the environment on a particular system can be summarized by its spectral density function, but other environments can produce equivalent effects. In both of these models, as well as in more general cases, the following conditions are essential for the appearance of dissipation (see, e.g. [Hu89]):

- (a) *System–environment separation.* This split depends on what one is interested in, which defines the system: it could be the slow variables, the low modes, the low order correlations, the mean fields; or what one is restricted to: the local domain, the late history, the low energy, the asymptotic region, outside the event horizon, inside the particle horizon, etc.
- (b) *Coupling.* The environment must have many degrees of freedom to share with and spread the information from the system; its coupling with the system

must be effective in the transfer of information (e.g. nonadiabatic) and the response of the coarse-grained environment must be sufficiently nonsystematic in that it will only react to the system in an incoherent and retarded way. (An example of almost the opposite condition is a dressed atom, i.e. an atom in a high finesse electromagnetic cavity where the quantum coherence of the system can be preserved to a high degree [CoPaPe95].)

- (c) *Coarse graining.* One must ignore or down-grade the full information in the environmental variables to see dissipation appearing in the dynamics of the open system. (The time of observation enters also, in that it has to be greater than the interaction time of the constituents but shorter than the recurrence time in the environment.) Coarse graining can be the causal truncation of a correlation hierarchy, the averaging of the higher modes, the “integrating out” of the fluctuation fields, or the tracing of a density matrix (discarding phase information).
- (d) *Initial conditions.* Whereas a dissipative system is generally less sensitive to the initial conditions in that for a wide range of initial states dissipation can drive the system to the same final (equilibrium) state, the process is nevertheless possible only if the initial state is off-equilibrium. The process manifests irreversibility also because the initial time is singled out as a special temporal reference point when the system is prepared in that particular initial state. Thus in this weaker sense, dissipation is also a consequence of specially prescribed initial conditions.¹

While the dynamics of the combined system made up of a subsystem and its environment is unitarity, and its entropy remains constant in time, when certain coarse graining is introduced in the environment, the subsystem turns into an open system, and the entropy of this open system (constructed from the reduced density matrix by tracing out the environmental variables) increases in time. In this open system dynamics, the effect of the coarse-grained environment on the subsystem leads to dissipation and irreversibility in its dynamics.

In our prior discussion of dynamical instability or “dispersion” with the example of hard-disk scattering we were introduced to irreversible but nondissipative processes. Irreversibility there refers to the ultra-sensitivity of the dynamics to the initial conditions. It is extremely difficult to trace back in time a highly divergent congruence of trajectories. The source of irreversibility

¹ Note the distinction between these cases: If one defines t_0 as the time when a dissipative dynamics begins and t_1 as when it ends, then the dynamics from t_0 to $-t$ is exactly the same as from t_0 to t , i.e. the system variable at $-t_1$ is the same as at t_1 . This is expected because of the special role assigned to t_0 in the dynamics with respect to which there is time-reversal invariance, but it is not what is usually meant by irreversibility in a dissipative dynamics. The arrow of time there is defined as the direction of increase of entropy and irreversibility refers to the inequivalence of the results obtained by reversing t_0 and t_1 (or, for that matter reversing t_0 and $-t_1$), but not between t_1 and $-t_1$. The time-reversal invariance of the H-theorem has the same meaning.

there is by nature fundamentally different from that found in open systems discussed here. The former dynamics is irreversible but nondissipative, while the latter is both dissipative and irreversible. Both types of processes depend on the stipulation of initial conditions. The difference is that the former depends sensitively so, the latter less sensitively. Thus dissipative processes must involve some measure of coarse graining, but coarse graining alone need not lead to dissipation. We will have a subsection later on the issue of coarse graining.

D. Phase mixing

Two well-known effects fall under this category: Landau damping and spin echo (e.g. [Bal75, Ma85]). Let us examine the first example. If one considers long-ranged forces such as the Coulomb force in a dilute plasma gas where close encounters and collisions are rare, the factorizable condition can be assumed to hold throughout, before and after each collision (thus there is no causal condition like the molecular chaos assumption imposed). Under these conditions the Boltzmann kinetic equation becomes a Vlasov (or collisionless Boltzmann) equation (see, e.g. [Bal75, Kre81]). This problem will be discussed in Chapters 10 and 11. The dependence on the one-particle distribution function $f_1(\mathbf{r}, \mathbf{p}, \mathbf{t})$ makes the Vlasov equation nonlinear, and it has to be solved in a self-consistent way. (This aspect is analogous to the Hartree approximation in many-body theory.) Note that the Vlasov equation which has a form depicting free streaming is time-reversal invariant: the Vlasov term representing the effect of the averaged field does not cause dissipation. This mean-field approximation in kinetic theory, which yields a unitary evolution of reversible dynamics, is, however, only valid for times short compared to the relaxation time of the system in its approach to equilibrium. This relaxation time is associated with the collision-induced dissipation process.

Landau damping in the collective local charge oscillations, being a solution of the Vlasov equation, is intrinsically a reversible process. The appearance of apparent “irreversibility” is a consequence of some specially stipulated initial conditions. One may even be able to find a function which is monotonically increasing and refer to it as representing entropy generation. However, upon the choice of some other condition, this feature can disappear and the entropy function can decrease. (An example in Chapter 4 is the entropy function defined in the particle number basis.) Landau “damping” is a mixing process, illustrated here by the Vlasov dynamics. It is fundamentally different from the dissipation process, in that the latter has an intrinsic damping time-scale but not the former, and that while dissipation depends only weakly on the initial conditions, mixing is very sensitive to the initial conditions. Spin echo is another well-known example of phase mixing [Bal75]. For quantum plasma, one needs to coarse grain the phase information in the wavefunctions *and* consider special initial conditions to see this apparent “damping” effect (more in Chapter 4).

From the array of examples above we see that irreversibility and dissipation involve very different causes. The effect of interaction, the role of coarse graining, the choice of time-scales, and the specification of initial conditions in any process can give rise to very different results. We will expand on these physical conditions later, after we have had a chance to look at the microscopic characterization of these macroscopic processes, i.e. their molecular dynamics basis.

1.2 Microscopic characterization from dynamical systems behavior

From a sampling of these macroscopic processes we see a variety of physical behavior. The underlying causes should all be traceable to the microscopic molecular dynamics, to which we now turn our attention. Let us start with a deceptively simple question: An isolated mechanical system is time-reversible. Under what conditions and in what sense does a large isolated system reach equilibrium?

1.2.1 Ergodicity describes a system in equilibrium

An isolated system of N molecules in a volume V has a constant total energy E under the Hamiltonian $H(\mathbf{r}, \mathbf{p})$, where \mathbf{r}, \mathbf{p} each is a $3N$ -dimensional vector denoting the position and momenta of all the particles in a $6N$ -dimensional phase space Γ . The density function $\rho(\gamma)$ is defined such that the probability of finding a member γ of the ensemble in a differential volume $d\Gamma \equiv d\mathbf{r}_1 \cdots d\mathbf{r}_N d\mathbf{p}_1 \cdots d\mathbf{p}_N$ is equal to $\rho(\gamma)d\Gamma$. Its dynamics is described by the **flow** of each member of the ensemble restricted to the constant energy surface or manifold \mathcal{E} in Γ . Since the number of members flowing in and out of a region in phase space should be equal for all times we have ρ satisfying the Liouville equation,

$$\frac{d\rho}{dt} \equiv \frac{\partial\rho}{\partial t} + \sum_1^N \left(\dot{\mathbf{r}}_i \cdot \frac{\partial}{\partial \mathbf{r}_i} + \dot{\mathbf{p}}_i \cdot \frac{\partial}{\partial \mathbf{p}_i} \right) \rho = 0 \quad (1.3)$$

where an overdot denotes derivative with respect to time.

In statistical mechanics the microcanonical **ensemble** describes such an isolated system. The number of states is represented by the area of the energy surface E in phase space:

$$\Omega(E) = \int_{H=E} d\mu \equiv \int_{\Gamma} \delta(H - E) d\mu, \quad (1.4)$$

where μ is the invariant measure on Γ . The entropy is defined as $S = k_B \ln \Omega(E)$. The ensemble average of a phase space function F over the energy surface E is given by

$$\langle F \rangle_{\mu} \equiv \frac{\int_{H=E} d\mu F(\gamma)}{\int_{H=E} d\mu} = \frac{\int_{\Gamma} F(\gamma) \delta(H - E) d\mu}{\int_{\Gamma} \delta(H - E) d\mu} \quad (1.5)$$

We also learned that a system in *equilibrium* (either by itself, as in a micro-canonical ensemble, or in contact with a heat bath, as in a canonical ensemble) will have *equal a priori probability* to occupy any of its accessible microstates. How do these concepts: ensemble average, flows in phase space and equilibrium state, connect with each other? Equivalence between the kinetic theory and statistical mechanics description implies that there must be a relation between the way the system points in phase space move (the Liouville flow) and what makes up a typical copy of the system (ensemble average). Equilibrium suggests that the system is stationary. Thus a typical system point must spend an equal amount of time in regions of phase space of equal measure on the energy surface. This is the gist of Boltzmann's **ergodic hypothesis**. If we define the time average of a phase space function $F(\gamma)$ on the energy surface E as

$$\langle F \rangle_t \equiv \lim_{t \rightarrow \infty} \frac{1}{T} \int_0^T F(\gamma_t) dt \quad (1.6)$$

where γ_t denotes the point in Γ space after evolving a time t , then the ergodic hypothesis states that

$$\langle F \rangle_t = \langle F \rangle_\mu. \quad (1.7)$$

This says that an arbitrary snapshot (time) of the system provides a typical copy (ensemble) of the system in equilibrium, or, loosely, that time average is equivalent to ensemble average.

Examples of an ergodic system include a one-dimensional harmonic oscillator, an automorphism on a 2-torus in phase space such as the baker's transform or the Arnold cat map. For a quantum system to be ergodic it has to have a nondegenerate energy spectrum. Many simple yet important systems in statistical mechanics are nonergodic. Examples are an ideal gas and multiple harmonic oscillators. For nonergodic systems the energy manifold is metrically decomposable, i.e. E can be partitioned into two or more invariant submanifolds each of which is invariant under the flow in phase space Γ . An equilibrium condition is described by an ensemble density which is constant on each submanifold, but not necessarily on the entire energy manifold.

Note that ergodicity is a microdynamics condition depicting a system in equilibrium but the property is irrelevant to whether a system can approach equilibrium [Far64].

1.2.2 *Mixing system is time-reversible; weak sense of approach to equilibrium*

One would think that if the flow in a system is chaotic enough such that the initial probability distribution spreads sufficiently evenly throughout the phase space then there may be a chance for the energy surface to be uniformly occupied. The first condition constitutes what is known as a **mixing system**. The second

condition is close to, but still insufficient to define a state of equilibrium. Let $\mu(A)$ be the measure² on a set A on the energy surface (the complete energy surface is denoted by \mathcal{E}) in phase space. Denote by A_t the same set at time t . It is obvious that $\mu(A) = \mu(A_t)$. A system is mixing if for all sets B on the energy surface the following holds:

$$\lim_{t \rightarrow \infty} \frac{\mu(A_t \cap B)}{\mu(B)} = \frac{\mu(A)}{\mu(\mathcal{E})} \quad (1.8)$$

In practice the infinite time limit can be just the laboratory or observation time-scale. What this means is that in a mixing system the stretching of the original set will enable it to intersect with almost any region in the entire energy surface. This requires two conditions. First, there must exist in the system trajectories which spread out rapidly in certain directions of the phase space on the energy surface. Second, that the flow can traverse the whole energy surface \mathcal{E} , so it has to be metrically nondecomposable (i.e. that it cannot be subdivided into two or more regions of nonzero measure such that a trajectory starting in one region will never leave it). Common examples of mixing systems are the baker's transform and the Arnold cat map. Since a mixing system has flows which are nearly uniform in the phase space it can be understood to imply ergodicity. However the converse is not true. Both ergodic and mixing systems are time-reversible.

To see irreversible behavior and the approach to equilibrium one needs to introduce some measure of coarse graining, such as considering only the slow variable associated with the unstable direction of the flows, or imposing certain assumptions on the initial conditions in the distributions.

1.2.3 Dissipative system: coarse-grained mixing permits approach to equilibrium

For a system whose unstable trajectories stretch out any initial distribution into very "long and narrow" filaments on the energy surface in the course of time they can produce a uniform spread in phase space. We refer to systems having these properties as satisfying the "chaotic hypothesis" of Cohen and Gallavotti [GalCoh95], or chaotic systems (strictly speaking, the system needs to satisfy the set of criteria which define a hyperbolic or an Anosov system, which is much stronger than mixing; to delve into this topic will go beyond the scope of our book, and we refer interested readers to nice monographs such as [Dor99, Gas98]). In such systems neighboring trajectories diverge from each other exponentially fast – with positive Liapunov exponents, in an unstable direction. The chaotic hypothesis is to dissipative systems as ergodicity is to equilibrium

² For hyperbolic systems, to capture the smoothness in the unstable directions and the fractal nature in the stable directions, one needs to use the Sinai–Ruelle–Bowen measure [RueEck85, Sinai72, BowRue75, Rue76].

systems. Thereupon one can speculate that averages taken with the distribution function defined on this extended set be equal to the average taken with a smooth equilibrium distribution. (To fulfill the exact criterion of equilibrium one also needs to consider whether the equal spreading in any region of phase space on the energy surface is also uniform in time.) Gibbs correctly observed that a mixing system will not reach the uniform phase space density $\bar{\rho}(\gamma)$ in the fine-grained sense, i.e. $\lim_{t \rightarrow \infty} \rho_t(\gamma) = \bar{\rho}(\gamma)$ for each phase space point γ in Γ . But it is likely to do so in a coarse-grained sense, i.e. that the average of $\rho_t(\gamma)$ over each fixed region of phase space will become uniform. $\bar{\rho}$ is called the weak limit of the family of functions ρ_t [Pen70].

It is only in this weak sense that a statement like “a mixing system approaches equilibrium” becomes valid. Bear in mind that a mixing system is time-reversal invariant without coarse graining. Even coarse graining does not automatically turn a mixing system into a dissipative one. We will expand on this point in the next section. Some additional conditions need to be introduced to turn a mixing system into one which shows irreversibility and approaches equilibrium.

Averaging and tracing. Since irreversibility appears in macroscopic systems one may attempt to scale up the system and hope that averaging over a larger phase space may lead to irreversibility. Coarse graining in this way does lead to entropy increase, but it does so in both time directions, so the system remains time-reversible. On the other hand, projection to a lower dimension, or “tracing” (in a sense defined by e.g. [Mac74]) does allow for increase in one time direction.

Molecular chaos assumption. Boltzmann’s *Stosszahlansatz*, or molecular chaos assumption, is a causal condition, i.e. before each collision the two molecules are uncorrelated but afterwards they are: This *is* where the irreversibility enters. The dynamical origin of Boltzmann’s *Stosszahlansatz* is not clear. Note the choice of time-scales involved here: the shorter time-scales describing higher order correlation functions play a less important role in the long-time behavior which is dominated by the time-scale associated with the behavior of the one-particle distribution function.

Unstable versus stable direction. On the microscopic level we notice already some coarse graining is introduced when we focus on the divergent trajectories in defining the unstable directions of flow in phase space. In the simplest example ([Dor99]) of a Boltzmann equation derived from the baker’s map it is the distribution function projected in the slow variables associated with the unstable direction which has a chance to approach the equilibrium distribution. This happens at a time-scale which is much shorter than the time it takes for any small region of phase space to get mixed in the full space. Again this presumes two conditions: that we are dealing with a projected distribution function by paying attention only to the trajectories with positive Lyapunov exponents, and

that the initial conditions are conducive to starting the chaotic flow. Notice the time factor involved. Here we see some similarity between our description of the molecular (micro) dynamics of a few degrees of freedom and the gas (macro) dynamics in terms of invoking projection (or reduction), imposing initial conditions, and choosing time-scales. Their connection certainly depends on the role of very large numbers, multi-time-scales or interaction strength. We now turn our attention to these issues.

1.2.4 Nonequilibrium thermodynamics and chaotic dynamics³

It is often said that information loss is the source of irreversibility and the approach to equilibrium. From the molecular dynamics description this is not always the case. The baker model gives a good example here. It is essential that, for chaotic systems at least (we really don't understand much about nonchaotic systems, paradoxically) the projection catches at least a piece of the unstable directions in phase space, so the stretching mechanism can smooth out an irregular distribution function. It is also important that the distribution functions be smooth enough not to be concentrated on special orbits which are insensitive to the projection. The application of the molecular chaos assumption in the Boltzmann theory is interesting in the sense that if one takes the correlations to be destroyed by collisions, then one gets an anti-Boltzmann equation, with funny properties. So to get irreversibility one needs both the projection onto a space that has some stretching mechanism in phase space as well as some special conditions imposed on an initial state distribution function.

The other form of coarse graining is more subtle and connected to chaos. Even if no projections onto lower dimensional phase spaces are made, chaotic dynamics, when present, forces a distribution function to become closer and closer to a fractal with structure on arbitrarily fine scales. In the limit of large times, distribution functions do not have nice mathematical properties. They are not differentiable, for example. Thus some coarse graining is required to go from an SRB fractal measure to a distribution function that can be used to calculate averages. This necessitates a loss of information and is a source of entropy increase.

The source of irreversibility in Landau damping is also connected with the construction of a fractal structure.⁴ Mathematically the distribution function becomes a Schwartz distribution and lives in a space where the usual theorems

³ The authors are grateful to Professor Robert Dorfman for sharing with them the latest view on these issues in a correspondence from which some of the description in this subsection is adopted.

⁴ In the opinion of some leading statistical physicists, Dorfman being one of them, who conveyed this to the authors, the earliest notion of fractals was introduced in the physics literature by van Kampen in his discussion of Landau damping. He called the distribution function "corrugated" for lack of a better word.

about the spectra of differential operators no longer apply, and decays can appear in unexpected ways. The notion of a “Gelfand triple” is useful here for its description [LaCaId99, ACGI00]. To delve further into these directions is beyond the scope and intent of this book, but interested readers should consult the excellent books of Dorfman [Dor99], Gaspard [Gas98] on the micro–macro relations. Pierre Gaspard has pioneered this approach to irreversibility and the Second Law by showing explicitly the deep connection between Kolmogorov–Sinai (KS) entropy and thermodynamic entropy production as well as other distinct properties of nonequilibrium thermodynamics from the theories of dynamical systems (see, e.g. the book of Nicolis [Nic95]). Some salient features are mentioned below (see Gaspard’s 2006 summer school lectures [Gas06]).

The aim is to understand the statistical behavior of a collection of particles such as relaxation, diffusion, dissipation, viscosity from the microdynamics of the particles and the divergence properties of their trajectories (congruence) in time. The starting point is the familiar Liouville equation for Hamiltonian systems. One can extract the intrinsic relaxation rates from this equation under certain assumptions on the dynamics. Two important quantities characterizing the microdynamics of the particle congruence are the Lyapunov exponents and the KS entropy per unit time. (The Lyapunov exponents characterize the sensitivity to initial conditions of the underlying microscopic dynamics while the KS entropy per unit time measures the degree of dynamical randomness developed by the trajectories of the system during their time evolution.) The new focus in this recent work is on the large deviations or large fluctuations that the dynamical properties of a system develop in time. In the escape-rate formalism, these large-deviation relationships relate these microscopic quantities to the transport functions in the macroscopic dynamics of the collective particles (see, e.g. [Gas98, Dor99]). These large-deviation relationships are also the basis for the formulation of new fluctuation theorems [EvCoMo93, EvaSea94, GalCoh95, Jar97]. The concepts and techniques in these interfaces have also proven to be invaluable in treating new mesoscopic physical phenomena at the nanoscale [BuLiRi05].

1.3 Physical conditions

Let us be reminded that in addition to examining the microscopic basis of nonequilibrium statistical mechanics via abstract dynamical systems we also need to consider the fact that we are dealing with a large system. It is well-known that the thermodynamic limit is obtained by taking N, V to infinity while keeping the ratio of these two quantities finite. What is the effect of a large system on fluctuations and irreversibility? How does the imposition of some specific initial condition alter the macroscopic dynamics? How can we understand the fundamental difference between microscopic and macroscopic behavior in terms of time-scales or interaction strength? How do the averaging or coarse-graining

procedures affect the outcome of a macroscopic observation? One needs to seek answers to these questions in order to address the fundamental issue of how the macroscopic features arise from its microscopic dynamics. We shall now combine the micro and macro descriptions in exploring these important physical factors. The following items also make up a useful checklist to examine whenever we encounter a new quantum field process and attempt to understand its basic statistical mechanical meaning.

1.3.1 Large systems: Fluctuations, Poincaré recurrence and thermodynamic limit

We are familiar with the advantage of taking the large number N and large volume V limit. *Thermodynamics* obtained in this limit while keeping $n = N/V$ constant is a simple yet powerful theory which captures the essential features of macroscopic phenomena. From microdynamics, a chaotic system (one which satisfies the “chaotic hypothesis”) approaches equilibrium in a coarse-grained sense ([Dor99]). For systems whose microdynamics has the right properties the average $\langle F \rangle_t$ of a dynamical variable F taken over the appropriate ensembles approaches an equilibrium value $\langle F \rangle_{Eq}$. To infer that in any of the individual systems in the ensemble F is close to $\langle F \rangle_{Eq}$ one needs to ensure that the fluctuations of F are small, and for this one needs to invoke the large size of the system as well as its mixing or chaotic properties ([Pen79]).

On the relation of fluctuations, the size of the system, and the time-scales involved, it is instructive to bring up the Poincaré recurrence and Zermelo’s (1896) objection to Boltzmann’s theory. Poincaré (1892) stated that any isolated, finite, conservative system will in a finite time come arbitrarily close to its initial configuration. Boltzmann’s $H_B(t)$ function cannot decrease monotonically but must eventually increase to reach its original value $H_B(0)$ in a finite time. Thus, Zermelo argued that Poincaré recurrence would undermine Boltzmann’s theory of approach to equilibrium.

Boltzmann’s answer to this paradox invokes **fluctuations** and **probabilistic** arguments. We know from statistics that if N were just a few particles the fluctuations are comparable to the mean. The Poincaré recurrence time T_P is short and there is no discernible trend of irreversibility. This case is not addressed by Boltzmann’s theory. The larger the system the smaller the fluctuations become, and the longer the Poincaré recurrence time. For example, Mazur and Montroll [MazMon70] considered a linear chain of N classical point masses m harmonically (with natural frequency ω_0) coupled to each other’s nearest neighbors. For $N = 10$ and $\omega_0 = 10/\text{sec}$ they found that $T_P = 10^{10}$ years, about the age of the universe. Only with a long Poincaré recurrence time will the distribution function for the macroscopic variables become sharp and the tenets of statistical mechanics apply. In addition to the size of the system the recurrence time depends sensitively and irregularly on the initial state. Because of random fluctuations

individual sample occurrences cannot be used for reliable prediction of the robust behavior of the overall physical systems, which can only be made in a probabilistic sense. Taking the thermodynamic limit permits one to construct a simpler, asymptotic, statistical theory for large systems. In this limit Poincaré's recurrence is probabilistically suppressed.

From kinetic theory considerations, the **dynamical correlations** established between particles after collisions will become less significant, at least for a dilute gas, when a larger system is considered by observers interested in the long-time behavior of the system. This enables one to focus on those physical quantities of most interest in the long-time limit, such as the expected value of the one-particle distribution function. It is in this same context where Boltzmann proposed his truly original and remarkable theory in depicting the dynamical behavior of the macroscopic world.

1.3.2 Initial conditions: Specific, randomized, dynamical correlations

For a mixing system any set of nonzero measure will be spread out in time uniformly on the energy surface. This suggests that the trajectories must be very sensitive to the initial conditions. Indeed it is so.

Boltzmann assumes that the molecular chaos assumption holds for each collision. Lanford in 1975 [Lan75], using a Lorenz gas of hard spheres of radius a , showed that in the (Grad) limit: $a \rightarrow 0$ while $n = N/V \rightarrow \infty$ in such a way that the mean free path $\lambda = (n\pi a^2)^{-1}$ remains constant (it thus applies for all values of the mean free path), and Boltzmann's *Stosszahlansatz* can be replaced by the assumption that the particles are uncorrelated initially: $\rho(z_1 \dots z_N; 0) = \prod_{i=1}^N f_i(z_i, 0)$ [$z_i = (\mathbf{r}_i, \mathbf{p}_i)$ denotes the coordinates and momenta of the i -th particle] since in this limit the r -particle distribution converges almost everywhere to products of one-particle distribution functions at all times, i.e. $\lim f_r(z_1, \dots, z_r, t) = \prod_{i=1}^r f_1(z_i, t)$. (Note that the Grad limit is different from the thermodynamic limit in that the volume is kept constant.)

Time-reversal of Boltzmann's dynamics can be exact but any small uncertainty or error can wipe out reversibility. Thus random initial conditions ensure that we can extract the system's generic and not specific behavior (by design, such as putting all particles on one side of a partition). Let us see how the **Loschmidt paradox** (1876) can illuminate the role played by the initial conditions on irreversibility. An isolated system is time-reversal invariant. If a system evolves towards equilibrium there must be an equally acceptable evolution which takes the system away from equilibrium which is not seen in nature. At the dawn of the computer age, one of the first computer simulations of molecular dynamics was performed by Orban and Bellemans in 1967 [OrbBel67, Ald73] who numerically integrated the equations of motion for a two-dimensional dilute gas (at a density of 0.04 of close packing) of 100 hard disks in a square box colliding

with each other and the box. They let the system evolve for a definite number of collisions up to time t_1 short compared to the equilibration time t_{Eq} , and then reversed all the velocities. Since this is a reversible microdynamics one may expect to recover the initial state after a time $2t_1$. They found that the accuracy with which the original state is restored at time $2t_1$ falls off rapidly as t_1 is increased, due to the rounding errors in the numerical integration. This can be understood in light of the divergence of trajectories (“dispersion,” or dynamical instability) discussed in Section 1.1. It is also a good illustration of the important role played by initial conditions: In the numerical simulation, Orban and Bellemans chose as *initial condition* $t = 0$ the molecules being placed at the vertices of a square mesh in the box with equal speed but *random direction*. The gas reaches equilibrium after about 200 collisions as the distribution approaches a Maxwell–Boltzmann form and the Boltzmann H-function of the velocities $H_B(t)$ reaches a minimum. In contrast the initial condition at t_1 for the time-reversed evolution is a very special one [Pri73], because the correlations established amongst the particles (hard disks) are very particular to that instant in the entire history of the system. If we consider the condition of the system close to equilibrium ($t \approx t_{\text{Eq}}$) as natural (highest probability of occurrence) then the condition of the system at t_1 is highly unnatural (very low probability) with respect to the equilibrium state. Indeed it shows anti-kinetic (contrary to Boltzmann’s predictions) behavior when $H_B(t)$ increases over a period of time. The result of these numerical simulations in spin-echo experiments was obtained by Rhim, Pines and Waugh in 1971 [RhPiWa71]. For experimental realization of the Loschmidt echo see [PLURH00].

Note also that the anti-kinetic behavior (Boltzmann’s H-function $H_B(t)$ increasing) cannot be obtained from solutions of the Boltzmann equation, because it is predicated upon a molecular chaos assumption. To do the velocity or time reversal one must solve for the correlations in time from the complete BBGKY hierarchy of the N -body system, which is difficult but possible numerically or experimentally, but almost impossible analytically.

Thus the resolution of the Loschmidt paradox is that Boltzmann’s equation is only an approximation to the exact equations of motion which describe systems with random initial states and no dynamical correlations. This is a very different situation from the time-reversed evolution, where the initial condition at t_1 registers information of strong **dynamical correlations**.

1.3.3 Time-scales and interaction

We have already seen how one can characterize the condition for a system to approach equilibrium by the discrepancy between characteristic time-scales. In Bogoliubov’s explanation of the kinetic conditions, Boltzmann’s equation governing the one-particle distribution function measures the time between collisions which is the slow variable (the relevant variable), while the fast variables giving

the time during collisions are ignored (the irrelevant variables). We saw a similar division of time-scales in the microdynamics of the chaotic systems. In such systems, the unstable direction defines a slow variable while a stable direction defines a fast variable. One can construct a Boltzmann equation (in the form of a gain–loss equation) which permits the approach to equilibrium (see [Dor99]). Such a Markovian equation shows irreversibility.

We also saw the relation of long time-scale (Poincaré recurrence) and large systems in relation to the formulation of thermodynamic and kinetic theories in the depiction of physical reality, likewise the dependence on the initial conditions.

It is often remarked that interaction (e.g. collisions amongst gas molecules) is needed for a system to equilibrate and to show irreversibility. Interaction is necessary for equilibration but interaction does not generate irreversibility. Mean field dynamics such as that described by the Vlasov equation has interaction but the dynamics is reversible. Equilibration (or thermalization when we refer to energy specifically but not particle numbers or chemical species) shows irreversibility, but irreversibility does not imply equilibration. We already saw at the molecular dynamics level that divergent trajectories show irreversibility, but it takes more to show equilibration (e.g. Anosov systems under coarse graining).

1.3.4 Coarse graining

Coarse graining in the most general sense refers to some information lost, removed, or degraded from a system. It could come about because this information is *inaccessible to us*, due to the limited accuracy in our observation or measurement. A drastic example is Planck-scale physics, the details of which are mostly lost (hard to retrieve) because the world we live in today is an ultra-low-energy construct. For this one needs to invoke ideas like effective field theory [Wei95]. Even when information is fully accessible to us in principle, in practice one may only be interested in some aspects of the system. *We choose to ignore* certain variables such as ignoring the higher order correlations in Boltzmann’s kinetic theory, or ignoring the phase information in a quantum system by imposing a random phase approximation. We do this by “integrating over” or “projecting out” these “irrelevant” variables.⁵

Let us see some examples of coarse graining in action. We start with the familiar Boltzmann theory: implementation of the molecular chaos assumption (i.e.

⁵ Quotation marks are put here to emphasize the colloquial usage and the warning that operations bearing the same name could bring forth different results depending on the assumptions introduced. For example, the projection operator formalism of Zwanzig and Mori *et al.* ([Nak58] [Zwa60] [Zwa61] [Mor65] [WilPic74] [Gra82] [Kam85] [GoKaZi04] [GorKar04] [Bal75]) applied to a closed system will turn the differential equations of motion for each subsystem into an integro-differential equation for a particular subsystem. Without casting away some information somewhere in the system this equation is just another way to express the interaction of the subsystems. It contains no more or no less information as the original equations describing the total system.

the two-particle distribution function $f_2 = f_1 f_1$ can be expressed schematically as a product of two one-particle distribution functions f_1) entails performing a coarse graining in the collision integral of space over the range of interaction and of time over the duration of a collision.

Another example concerns particle creation from an external background field or changing spacetime. Since particle pairs originate from the vacuum which is a pure state, there should not be any entropy generation. On the other hand, in a thermodynamics description, the entropy S is related to the number N of particles present. We may wonder whether to trust either or both of these statements. The key lies in understanding that the thermodynamic description has undergone several levels of coarse graining from the fundamental quantum field theory description. Indeed it is a very educational intellectual exercise to see what coarse-graining measures are introduced and what concepts are at work as we move from a microscopic (quantum field theory) to a macroscopic (nonequilibrium thermodynamics) description of this same system, but with different degrees of precision.

Note also that coarse graining is a necessary but not sufficient condition for entropy generation. It does not always produce a dissipative system. Truncation of the BBGKY hierarchy leads to a closed subsystem composed of n -particle correlation functions whose dynamical equations are unitary. (An example mentioned before is the Vlasov equation describing particle interaction via long-range forces.) In quantum field theory equations derived from a finite-loop effective action are also unitary – at one loop the effect of the quantum field on the particles manifests through the renormalized masses and charges (to be exact, the equations of motion derived from a finite-loop effective action are unitary if none of the relevant correlation functions are “slaved” – see Chapter 6 Section 6.3 and Chapter 9 Section 9.2.3 for a discussion of this concept; for ℓ loops, one must keep the first $(\ell + 1)$ th-order correlations, otherwise dissipation in the sense defined above sets in – dissipation is absent only in very specific situations, such as a free theory or equilibrium initial conditions). That is perhaps why (if one limits one’s attention to loop expansions) statistical mechanical concepts rarely came to the fore, until one starts asking questions of a distinct nature, such as how dissipative dynamics appears in an otherwise unitary system, and the origin and nature of noise in quantum field theory. A causal condition needs to be introduced to render the dynamics of the subsystem irreversible. This opens up another important theme in this book: effective field theory viewed in the open system framework, which will be developed in later chapters.

1.4 Coarse graining and persistent structure in the physical world

We have seen from the above discussions that the appearance of irreversibility is often traced to the initial condition being special in some sense. The dynamics of the system and how it interacts with its environment also enter

in determining whether the system exhibits mixing or dissipative behavior. For the sake of highlighting the contrast we could broadly divide the processes into two classes depending on how sensitive they are to the initial conditions versus the dynamics.⁶ One can say that the first class is *a priori* determined by the initial conditions, the other is *a posteriori* rather insensitive to the initial conditions. Of the examples we have seen, the first group includes divergent trajectories in molecular (micro) dynamics, Landau damping, vacuum particle creation, and the second class includes gas (macro) or fluid dynamics (see the discussions at the end of Section 1.2.3), diffusion, particle creation with interaction, decoherence. Appearance of dissipation is accompanied by a degradation of information via coarse graining (such as the molecular chaos assumption in kinetic theory, restriction to one-particle distribution in particle creation with interaction, “integrating out” some class of histories in decoherence). An arrow of time appears because of some special prearranged conditions; how it manifests in the system also depends on the system dynamics and the coarse graining introduced to the system. The issues we have touched on involve the transformation of a closed to an open system, the relation between the microscopic and the macroscopic world, and the transition from quantum multiplicities to classical realities. Many perceived phenomena in the observable physical world, including the phenomenon of time-asymmetry, can be understood in the open-system viewpoint via the approximations introduced to the objective microscopic world by a macroscopic observer [GKJKSZ96, Omn94, Per93].

Thus, time asymmetry in these processes is influenced by many factors: the way one stipulates the boundary conditions and initial states, the time-scale of observation in comparison with the dynamical time-scale, how one decides what the relevant variables are and how they are separated from the irrelevant ones, how the irrelevant variables are coarse grained, and what assumptions one makes and what limits one takes in shaping a macroscopic picture from one’s imperfect knowledge of the underlying microscopic structure and dynamics.

We have discussed the procedures which can bring about these results. However, a set of more important and challenging issues remain largely unexplored, i.e. under what conditions the outcomes become less subjective and less sensitive to these procedures, such as the system–environment split and the coarse graining of the environment. These procedures provide one with a viable prescription to get certain general qualitative results, but are still not specific and robust enough to explain how and why the variety of observed phenomena in the physical world arise and stay in their particular ways. To address these issues one should ask a different set of questions:

⁶ As discussed earlier, dissipation also requires the stipulation of a somewhat special initial condition, i.e. that the system is not in an equilibrium state; but, in the words of R. Sorkin, “not more special than it needs to be”.

(1) *By what criteria are the system variables chosen? Collectivity and hierarchy of structure and interactions*

In a model problem, one picks out the system variables – be it the Brownian particle or the mini-superspace variables – by fiat. One defines one's system in a particular way because one wants to calculate the properties of that particular system. But in the real world, certain variables distinguish themselves from others because they possess a relatively well-defined, stable, and meaningful set of properties for which the observer can carry out measurements and derive meaningful results. Its meaningfulness is defined by the range of validity or degree of precision or the level of relevance to what the observer chooses to extract information from. In this sense, it clearly carries a certain degree of subjectivity – not in the sense of arbitrariness in the exercise of free will of the observer, but in the specification of the parameters of observation and measurement. For example, the thermodynamic and hydrodynamic variables are only good for systems close to equilibrium; in other regimes one needs to describe the system in terms of kinetic-theoretical or statistical-mechanical variables.

The soundness in the choice of a system in this example thus depends on the time-scale of measurement compared to the relaxation time. As another example, contrast the variables used in the nuclear collective model and the independent nucleon models. One can use the rotational–vibrational degrees of freedom to depict some macroscopic properties of the motion of the nucleus, and one can carry out meaningful calculations of the dissipation of the collective trajectories (in the phase space of the nucleons) due to stochastic forces. In such cases, the noncollective degrees of freedom can be taken as the noise source. However, if one is interested in how the independent nucleons contribute to the properties of the nucleus, such as the shell structure, one's system variable should, barring some simple cases, not be the elements of the $SO(3)$ group, or the $SU(6)$ group. At a still higher energy where the attributes of the quarks and the gluons become apparent, the system variables for the calculation of, say, the stability of the quark–gluon plasma should change accordingly. The level of relevance which defines one's system changes with the level of structure of matter and the relative importance of the forces at work at that level. The improvement of the Weinberg–Salam model with W, Z intermediate bosons over the Fermi model of four-point interactions is what is needed in probing a deeper level of interaction and structure which puts the electromagnetic and weak forces on the same footing. Therefore, one needs to explore the rules for the formation of such relatively distinct and stable levels, before one can sensibly define one's system (and the environment) to carry out meaningful inquiries of a statistical nature.

What is interesting here is that these levels of structures and interactions come in approximate hierarchical order (so one doesn't need QCD to calculate the rotational spectrum of a nucleus, and the Einstein spacetime manifold picture

will hopefully provide most of what we need in the post-Planckian era). One needs both some knowledge of the hierarchy of interactions and the way effective theories emerge from “integrating out” variables at very different energy scales in the hierarchical structure (e.g. ordinary gravity plus particle theory regarded as a low-energy effective higher-dimension or Kaluza–Klein theory). The first part involves fundamental constituents and interactions and the second part the application of statistical methods. One should also keep in mind that what is viewed as fundamental at one level can be a composite or statistical mixture at a finer level. There are system–environment separation schemes which are designed to accommodate or reflect these more intricate structures, from the mean-field–fluctuation-field split to the multiple source or nPI formalism (see Chapter 6) for the description of the dynamics of correlations and fluctuations. The validity of these approximations depends quite sensitively on where exactly one wants to probe in between any two levels of structure. Statistical properties of the system such as the appearance of dissipative effects and the associated irreversibility character of the dynamics in an open system certainly depend on this separation.

(2) How does the behavior of the subsystem depend on coarse graining? Sensitivity and variability of coarse graining, stability and robustness of emergent structure

Does there exist a common asymptotic regime as the result of including successively higher order iterations in the same coarse-graining routine? This measures the sensitivity of the end result to a particular kind of coarse graining. How well can different kinds of coarse-graining measure produce and preserve the same result? This is measured by its variability. Based on these properties of coarse graining, one can discuss the relative stability of the behavior of the resultant open system after a sequence of coarse grainings within the same routine, and its robustness with respect to changes to slightly different coarse-graining routines.

Let us illustrate this point with some simple examples. When we present a microscopic derivation of the transport coefficients (viscosity, heat conductivity, etc.) in kinetic theory via the system–environment separation scheme, we usually get the same correct answer independent of the way the environment is chosen or coarse grained. Why? It turns out that this is likely only if we operate in the linear-response regime (see [FeyVer63]). The linear coupling between the system and the environment makes this dependence simple. This is something we usually take for granted, but has some deeper meaning. For nonlinear coupling, the above problem becomes nontrivial. Another aspect of this problem can be brought out [BalVen87, Spo91] by comparing these two levels of structure and interaction, e.g. the hydrodynamic regime and the kinetic regime. Construct the relevant entropy from the one-particle classical distribution function f_1 , that gives us the kinetic theory entropy S_{kt} which is simply $-kH_{\text{B}}$, where H_{B} is Boltzmann’s H-function. Now comparing it with the hydrodynamic entropy function S_{hd} given

in terms of the hydrodynamic variables (in this case, the number and energy density), one sees that $S_{\text{hd}} > S_{\text{kt}}$. A simple physical argument for this result is that the information contained in the correlations amongst the particles is not included in the hydrodynamic approximation. Even within the kinetic theory regime there exist intermediate stages described by suitably chosen variables [Spo91]. The entropy functions constructed therefrom will reflect how much fine-grained information is lost. In this sense S_{hd} is a maximum in the sequence of different coarse-graining procedures. In the terminology we introduced above, by comparison with the other regimes, the hydrodynamic regime is more robust in its structure and interactions with respect to varying levels of coarse graining. One way to account for this is that, as we know, the hydrodynamic variables enter in the description of systems in equilibrium and they obey conservation laws [HaLaMa95, Bru96, Hal98]. Further coarse graining on these systems is expected to produce the same results, i.e. the hydrodynamic regime is a limit point of sorts after the action from a sequence of coarse grainings. Therefore, a kind of “maximal entropy principle” with respect to variability of coarse graining is one way where thermodynamically robust systems can be located.

While including successively higher orders of the same coarse-graining measure usually gives rise to quantitative differences (if there is a convergent result, that is, but this condition is not guaranteed, especially if a phase transition intervenes), coarse graining of a different nature will in general result in very different behavior in the dynamics of the open system. Let us look further at the relation of variability of coarse graining and robustness of structure.

Sometimes the stability of a system with respect to the variability of coarse graining is an implicit criterion behind the proper identification of a system. For example, Boltzmann’s equation governing the one-particle distribution function which gives a very adequate depiction of the physical world is, as we have seen, only the lowest order equation in an infinite (BBGKY) hierarchy. If coarse graining is by the order of the hierarchy – e.g. if the second and higher order correlations are ignored – then one can calculate without ambiguity the error introduced by such a truncation. The dynamics of the open system which includes dissipation effects and irreversible behavior will not change drastically if one uses a different (say more fine-grained) procedure, such as retaining the fourth-order correlations (if the series converges, which is a nontrivial issue, see, e.g. [Dor81]). Consider now a different approximation: For a binary gas of large mass discrepancy, if one considers the system as the heavy mass particles, ignore their mutual interactions and coarse grain the effect of the light molecules on the heavy ones, the system now behaves like a Brownian particle motion described by a Fokker–Planck equation. We get a qualitatively very different result in the behavior of the system.

In general the variability of different coarse grainings in producing a qualitatively similar result is higher (more variations allowed) when the system one works with is closer to a stable level in the interaction range or in the hierarchical

order of structure of matter. The result is more sensitive to different coarse-graining measures if it is far away from a stable structure, usually falling in between two stable levels.

One tentative analogy may help to fix these concepts. Robust systems are like the stable fixed points in a parameter space in the renormalization group theory description of critical phenomena: the points in a trajectory are the results of performing successive orders of the same coarse-graining routine on the system (e.g. the Kadanoff–Migdal scaling [Kad76, Kad77, WilKog74, Fis74, Fis83]), a trajectory will form if the coarse-graining routine is stable. An unstable routine will produce in the most radical situations a random set of points. Different trajectories arise from different coarse-graining routines. Neighboring trajectories will converge if the system is robust, diverge if not. Therefore the existence of a stable fixed point where trajectories converge to is an indication that the system is robust. Only robust systems survive in nature and carry definite meaning in terms of their persistent structure and systematic evolution. This is where the relation of coarse graining and persistent structures enters.

So far we have only discussed the activity around one level of robust structure. To investigate the domain lying in-between two levels of structures (e.g. between nucleons and quark–gluons) one needs to first know the basic constituents and interactions of the two levels. This brings back our consideration of levels of structures above. Studies in the properties of coarse graining can provide a useful guide to venture into the often nebulous and elusive area between the two levels and extract meaningful results pertaining to the collective behavior of the underlying structure. But one probably cannot gain new information about the fine structure and the new interactions from the old just by these statistical measures (cf. the old bootstrapping idea in particle physics versus the quark model).⁷

1.5 Physical systems: Closed, open, effectively closed and effectively open

1.5.1 *Open systems: Coarse graining and back-reaction*

In treating physical systems containing many degrees of freedom one often attempts to select out a small set of variables to render the problem technically tractable while preserving its physical essence. Familiar examples abound: e.g. thermodynamics from statistical mechanics, hydrodynamic limit of kinetic theory, collective dynamics in condensed matter and nuclear physics [Wil82].

⁷ In this sense, one should not expect to gain new fundamental information about quantum gravity just by extrapolating what we know about the semiclassical theory, although studying the way the semiclassical theory takes shape (viewed as an effective theory) from possible more basic theories is useful. It may also be sufficient for what we can understand or care about in this later stage of the universe we now live in.

When one starts from the microscopic picture, one distinguishes the variables which depict the system of interest from those which can affect the system but whose detail is otherwise of lesser interest or importance. Making a sensible distinction involves recognizing and devising a set of criteria to separate the relevant from the irrelevant variables. This procedure is simplified when the two sets of variables possess very different characteristic time or length or energy scales or interaction strengths. An example is the separation of slow-fast variables as in the Born–Oppenheimer approximation in molecular physics where the nuclear variables are assumed to enter adiabatically as parameters in the electronic wavefunction. Similar separation is possible in quantum cosmology between the “heavy” gravitational sector characterized by the Planck mass and the “light” matter sector. In statistical physics this separation can be made formally with projection operator techniques. This usually results in a nonlinear integro-differential equation for the relevant variables, which contains the causal and correlational information from their interaction with the irrelevant variables.

Apart from finding some way of *separating* the overall closed system into a “relevant” part of primary interest (the open system) and an “irrelevant” part of secondary interest (the environment) in order to render calculations possible, one also needs to devise some *averaging* scheme to reduce or reconstitute the detailed information of the environment such that its effect on the system can be represented by some macroscopic functions, such as the transport functions. This involves introducing certain *coarse-graining* measures. It is usually by the imposition of such measures that an environment is turned into a bath, and certain macroscopic characteristics such as temperature and chemical potential can be introduced to simplify its description. A coarse-grained description of the effect of the environment on the system (in terms of, say, thermodynamic or hydrodynamic variables and their associated response functions) is qualitatively very different from the detailed description (in terms of the underlying microscopic variables and dynamics). A familiar example in many-body theory used for simplifying the effect of the environment is by assuming that each independent particle interacts with an effective potential depicting the averaged effect of all other particles. Vlasov dynamics in a plasma is of such a nature, so is the mean field approximation in quantum field theory (where the effect of quantum fluctuations of fields is described at this level of approximation in terms of a renormalized interaction potential and couplings to the system).

How good an *effective theory* is in its depiction of physical phenomena at a particular scale is usually determined by the appropriateness in the choice of the collective variables, the correctness and extent of coarse graining in relation to the probing scale and the precision of measurement. How the environment affects the open system is determined by the **back-reaction** effects. By referring to an effect as a back-reaction, it is implicitly assumed that a system of interest is preferentially identified, that one cares much less about the details of the other sector (the “irrelevant” variables in the “environment”). The

back-reaction can be significant, but should not be too overpowering, so as to invalidate the separation scheme. To what extent one views the interplay of the two sectors as *interaction* (between two subsystems of approximately equal weight) or as *back-reaction* (of a less relevant environment on the more important system) is reflexive of and determined by the degree to which one decides to keep or discard the information in one subsystem versus the other. It also depends on their interaction strength. Through reaction and back-reaction the behavior of each sector is linked to the other in an inseparable way, i.e. by their interplay.

Self-consistency is thus a necessary requirement in back-reaction considerations. This condition can manifest itself as the fluctuation–dissipation theorem (FDT). When the environment is a bath, for systems near equilibrium, their response can be depicted by linear response theory. Even though such relations are usually presented in such a context, its existence in a more general form can be shown to cover nonequilibrium systems. Indeed as long as back-reaction is included, such a relation can be understood as a corollary of the self-consistency requirement, which ultimately can be traced to the unitary condition of the original closed system.

A familiar example of a self-consistent back-reaction process is the time-dependent Hartree–Fock approximation in atomic physics or nuclear physics, where the system could be described by the wavefunction of the electrons obeying the Schrödinger equations with a potential determined by the charge density of the electrons themselves via the Poisson equation. In a cosmological back-reaction problem, one can view [Hu89] the system as a classical spacetime, whose dynamics is determined by Einstein’s equations with sources given by particles produced by the vacuum excited by the dynamics of spacetime and depicted by the appropriate wave equations in this particular curved spacetime [BirDav82].

Much of the physics of open systems is concerned with the appropriateness in the devising and the implementation of these procedures. They are: (1) the identification and separation of the physically interesting variables which make up the open system – one needs to first come up with the appropriate collective variables; (2) the “averaging” away of the environment or irrelevant variables – how different coarse-graining measures affect the final result is important (as discussed in the last section); and (3) the evaluation of the averaged effect of the environment on the system of interest. We will refer to these procedures as *separation*, *coarse graining* and *back-reaction* for short.

These considerations surrounding an open system are common and essential not only to well-posed and well-studied examples of many-body systems like molecular, nuclear and condensed-matter physics, they also bear on some basic issues at the foundations of quantum mechanics and statistical mechanics, such as decoherence and the existence of the classical limit [HarGel93], with profound implications on the emergence of time and spacetime [Har92], or quantum mechanics itself [Adl04].

1.5.2 From closed to effectively open systems

There are many systems in nature which are apparently closed (to the observer), in that there is no obvious way to define a system which is so much different from an environment. These systems do not possess a parameter which can enable the observer to distinguish possible heavy–light sectors, high–low frequency behavior or slow–fast dynamics. Boltzmann’s theory of molecular gas is a simple good example: All molecules in the gas are on an equal footing, in that no one can claim to be more special than the others. Because of the lack of parameters which marks the discrepancy of one component from the other, these systems do not lend themselves to an obvious or explicit separation from their environment (like open systems would), and appear like closed systems. However, usually in their effective description a separation is introduced implicitly or operationally because of their restricted appearance or due to the imprecision in one’s measurement. These are called effectively open systems.

In this example, on the microscopic level (of molecular dynamics) all molecular movements are time-reversal invariant, but on the macroscopic level (of our observation), dissipation and violation of time-reversal invariance obviously exist. To reconcile this difference and understand the origin of dissipation in nature was of course the great challenge Boltzmann posed for himself and which he so ingeniously resolved. Boltzmann came up with the idea that if only one-particle distribution functions were observed, and the molecular chaos assumption was imposed (for any collision process), there is an explanation for the origin of dissipation in macroscopic phenomena. Using the correlation functions (the lowest order being the one-particle distribution function) as a way to systemize the information in the gas, one would get the BBGKY hierarchy, which contains the full information of the gas. It is only upon the truncation of the hierarchy and the re-expression of the higher correlation functions in terms of the lower ones, e.g. the causal factorization condition (assuming that colliding partners are uncorrelated initially, what we shall call “slaving,” to be discussed in detail later), that the otherwise closed system expressed by the full hierarchy is rendered open, and dissipation appears. It is in this sense that it is called an effectively open system.

1.5.3 Two major paradigms of nonequilibrium statistical mechanics

We can highlight the distinction between open and effectively open systems by comparing the two primary models which characterize these two major paradigms of nonequilibrium statistical mechanics (see, e.g. [AkhPel81, Pri62, ToKuSa92]): the Boltzmann–BBGKY theory of molecular kinetics, and the Langevin (Einstein–Smoluchowski) theory of Brownian motions. The differences between the two are of both formal and conceptual nature.

To begin with, the *setup* of the problem is different: As we remarked above, in kinetic theory one studies the overall dynamics of a system of gas molecules,

treating each molecule in the system on the same footing, while in Brownian motion one (Brownian) particle which defines the system is distinct, the rest are relegated as the environment. The terminology of “relevant” versus “irrelevant” variables not so subtly reflects the discrepancy.

The *object* of interest in kinetic theory is the (one-particle) distribution function (or the n th-order correlation function), while in Brownian motion it is the reduced density matrix. The emphasis in the former is the behavior of the gas as a whole (e.g. dissipative dynamics) taking into account the correlations amongst the particles, while in the latter it is the motion of the Brownian particle under the influence of the environment.

The nature of *coarse graining* is also very different: in kinetic theory coarse graining resides in confining one’s attention to one-particle distribution functions, a factorization condition for the two-point functions and the adoption of the molecular chaos assumption. This corresponds formally to a truncation of the BBGKY hierarchy and introducing a causal slaving condition, while in Brownian motion it is in the integration over the environmental variables. The part that is truncated or “ignored” is where the noise comes from, while its main physical effect on the “system” is to render its dynamics dissipative. Thus the fluctuation–dissipation relation and other features.

Finally the *philosophies* behind these two paradigms are quite different: In Brownian motion problems, the separation of the system from the environment is prescribed: it is usually determined by some clear disparity between the two systems. These models represent “autocratic systems,” where some degrees of freedom are more relevant than others. By contrast, molecular gas models subscribing to the effectively open systems represent “democratic systems” where all particles in a gas are equally relevant. In the lack of any clear discrepancy in scales, making a separation “by hand” is *ad hoc*, contrived, and often leading to wrong description. Coarse graining in Boltzmann’s kinetic theory is also very different from that of Brownian motion. The latter is explicit while the former appears implicit (having its own systematics). However, as we shall see later, the coarse graining in the Boltzmann theory lies in the truncation and slaving procedures, where information attached to higher correlation orders is not kept in full. Now just what correlation order is sufficient for the physics under study is an objectively definable and verifiable fact, which ultimately is determined by the degree of precision in a measurement and judged by how well it depicts the relevant physics.

In Chapter 2 we will provide a physical discussion of the Boltzmann and Langevin dynamics, two prime examples of these two major paradigms. To see the mathematical origin of these stochastic equations it is best to acquire some knowledge of stochastic processes. A brief summary of this subject is given in Appendix A, which starts with rudimentary probability theory and ends with a derivation of the Chapman–Komogorov/Einstein–Smolochousky equation and the Kramers–Moyal/Fokker–Planck equation.

1.6 Appendix A: Stochastic processes and equations in a (tiny) nutshell

We give here a brief summary of the theory of stochastic processes, leading to the derivation of the Chapman–Komogorov/Einstein–Smolochousky equation and the Kramers–Moyal/Fokker–Planck equation. We will convey the necessary yet minimal set of information to enable the reader without prior knowledge of this subject to follow the development of later chapters leading to its application to quantum field theory of nonequilibrium processes. The emphasis here is more on physical ideas than mathematical rigor. One can find nice discussions of these topics in standard books such as those by van Kampen [Kam81], Papoulis [Pap84], Gardiner [Gar90], Gardiner and Zoller [GarZol00b], Carmichael [Car93], and Reichl [Rei98]. More succinct and accessible summaries can be found in, for example, Weissbluth [Wei88] and Mandel and Wolf [Man95]. Here we follow mainly the discussions by van Kampen.

1.6.1 Probability, random variables and stochastic processes

Probability

We start with some basic concepts and definitions in *probability theory*. A probability space, or simply, an experiment, consists of the triplet (Ω, F, P) where Ω is the sample space containing all possible outcomes of the experiment. An event A is a subset of Ω , and F is a collection of subsets of Ω . P is the probability of finding A in such an experiment. Example: In a single throw of a dice (our experiment) what is the probability of finding an even number facing up? Then $\Omega = \{1, 2, 3, 4, 5, 6\}$, $A = \{2, 4, 6\}$, $P = \frac{1}{2}$. Set theory is usefully applied to probability theory starting with $P(\Omega) = 1$, $P(\emptyset) = 0$, where \emptyset denotes the empty set. Two events are said to be **mutually exclusive**, or disjoint, if $P(A \cup B) = P(A) + P(B)$ or $P(A \cap B) = 0$. Two events A, B are **independent** iff $P(A \cap B) = P(A)P(B)$. Note that independent events are not mutually exclusive events because for mutually exclusive events $P(A \cap B) = 0$.

Let A_1, \dots, A_n be a finite collection of events. They are called mutually independent if for any $1 \leq i_1 < i_2 < \dots < i_k \leq n$

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2}) \dots P(A_{i_k}) \quad (1.9)$$

The events are called pairwise independent if for any $1 \leq i_1 < i_2 \leq n$

$$P(A_{i_1} \cap A_{i_2}) = P(A_{i_1})P(A_{i_2}) \quad (1.10)$$

Obviously pairwise independence is a much weaker condition than mutual independence.

Finally we define the conditional probability $P(A|B)$ as the probability that event A will occur if B occurs, or simply, A given B. Obviously the probability

that both events A, B will occur is $P(A \cap B) = P(A|B)P(B)$. From this it is easy to derive Bayes' rule

$$P(A|B)P(B) = P(B|A)P(A) \quad (1.11)$$

Random variables

A **random variable** X defined on a sample space Ω is a function which maps Ω into the set of real numbers. It assigns a real number to each sample point. In the example of throwing a dice, winning a certain amount of money x_i (out of the whole range $X(\Omega) = \{x_i\}$) when some number in the set $i = \{1 \cdots 6\}$ faces up is one such mapping. One can define a probability distribution $P(x_i) = f(x_i)$. For continuous variables the probability P that an event occurs resulting in X taking on values in an interval $a \leq X \leq b$ is given by

$$P(a \leq X \leq b) = \int_a^b f_X(x) dx \quad (1.12)$$

$f_X(x)$ is called the probability density for such an occurrence. The cumulative probability distribution function (cdf) $F_X(x) \equiv P(X \leq x)$ is obtained by integrating f_X up to the value x , i.e.

$$F_X(x) = \int_{-\infty}^x f_X(x') dx' \quad (1.13)$$

Now consider two stochastic variables X, Y on the same sample space Ω , $X(\Omega) = \{x_i\}, Y(\Omega) = \{y_j\}$. We define the **joint probability distribution** $f(X, Y)$ of X and Y as the probability of an ordered pair occurring, $P(x_i \in X, y_j \in Y) = f(x_i, y_j)$. For continuous variables $f(x, y) \geq 0$ is normalized $\int \int dx dy f(x, y) = 1$. The single-variable distribution f_X is obtained if one disregards (integrates over) the value of Y in $f(x, y)$, i.e.

$$f_X(x) = \int dy f(x, y), \quad f_Y(y) = \int dx f(x, y) \quad (1.14)$$

We now generalize the number of stochastic variables to r and define an r -dimensional vector $\mathbf{X} = (X_1, X_2, \dots, X_r)$. We can think of this as the vector denoting the space and momenta of N particles in the phase space Γ , in which case $r = 6N$. (Note we used i, j earlier to denote the (discrete) sample space variables, while r, s here denote the dimension of the space of stochastic variables.) The probability density $P_r(\mathbf{X})$ is the joint probability density of the r variables (X_1, X_2, \dots, X_r) . For a projection of \mathbf{X} into a smaller space of dimension s , $s < r$, the joint probability density of a subset $s < r$ of variables $P_s(X_1, X_2, \dots, X_s)$ regardless of the remaining variables X_{s+1}, \dots, X_r is obtained from integrating over these variables, i.e.

$$P_s(X_1, X_2, \dots, X_s) = \int P_r(X_1, \dots, X_s, X_{s+1}, \dots, X_r) dX_{s+1} \cdots dX_r \quad (1.15)$$

In probability theory this is called the **marginal distribution** of subset r . In statistical mechanics this is called the **reduced (probability density) distribution function**.

One can define the **n th moments** of a stochastic variable X by

$$\langle X^n \rangle = \sum_i x_i^n f_X(x_i) = \int dx x^n f_X(x) \quad (1.16)$$

where the first is for discrete and the second for continuous variables. The first two moments are familiar: For $n = 1$, $\langle X \rangle$ is called the **mean**; for $n = 2$, $\sigma_X^2 = \langle X^2 \rangle - \langle X \rangle^2$ is called the **variance**.

One can also define the **characteristic function** Φ as the Fourier transform of the probability density

$$\Phi_X(k) = \langle e^{ikX} \rangle = \int_{-\infty}^{+\infty} dx e^{ikx} f_X(x) = \sum_{n=0}^{\infty} \frac{(ik)^n}{n!} \langle X^n \rangle \quad (1.17)$$

with inverse transform

$$f_X(x) = \frac{1}{2\pi} \int dk e^{-ikx} \Phi_X(k) \quad (1.18)$$

Equivalently, in terms of Laplace transforms we can define the **moment generating function**

$$M_X(s) = \langle e^{sX} \rangle \quad (1.19)$$

This name becomes obvious when we rewrite the **moments** as

$$\langle X^n \rangle = \left. \frac{d^n M_X(s)}{ds^n} \right|_{s=0} \quad (1.20)$$

Likewise one can define a **cumulant** expansion by the relation

$$\ln \Phi_X(k) = \sum_{n=1}^{\infty} \frac{(ik)^n}{n!} C_n(X) \quad (1.21)$$

The relations between cumulants and moments are as follows:

$$\begin{aligned} C_1(X) &= \langle X \rangle & C_2(X) &= \langle X^2 \rangle - \langle X \rangle^2 \\ C_3(X) &= \langle X^3 \rangle - 3 \langle X^2 \rangle \langle X \rangle + 2 \langle X \rangle^3 \\ C_4(X) &= \langle X^4 \rangle - 4 \langle X^3 \rangle \langle X \rangle - 3 \langle X^2 \rangle^2 \\ &\quad + 12 \langle X^2 \rangle \langle X \rangle^2 - 6 \langle X \rangle^4 \end{aligned} \quad (1.22)$$

Note again that the first two cumulants are the mean and the variance. The **covariance** and **correlation** of two different stochastic variables X, Y are

defined respectively as

$$\begin{aligned}\text{Cov}(X, Y) &\equiv \int \int dx dy (x - \langle X \rangle)(y - \langle Y \rangle) f(x, y) \\ &= \langle XY \rangle - \langle X \rangle \langle Y \rangle \\ \text{Cor}(X, Y) &\equiv \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}\end{aligned}\quad (1.23)$$

Stochastic processes

Given a stochastic variable X , one can define a **stochastic function** Ξ obtained from X by some mapping

$$\Xi_X(t) = g(X, t) \quad (1.24)$$

where t is some smooth variable. If t denotes time, $\Xi(t)$ is called a **stochastic process**. When X takes on the value x , $\Xi_x(t) = g(x, t)$ becomes a sample function or a realization of the process.

The probability density for a stochastic function $\Xi_x(t)$ to take on value ξ at time t is given by

$$P_1(\xi, t) = \int \delta(\xi - \Xi_x(t)) f_X(x) dx \quad (1.25)$$

We recognize that $f_X(x)$ is the probability density for the stochastic variable X . The subscript 1 denotes a function of one stochastic variable. The probability over all values of ξ_1 at any particular time t_1 should be unity, thus the normalization condition is $\int P_1(\xi, t_1) d\xi = 1$. Generalizing to n we can define the **joint probability density** as

$$P_n(\xi_1, t_1, \dots, \xi_n, t_n) \equiv \int \delta(\xi_1 - \Xi_x(t_1)) \cdots \delta(\xi_n - \Xi_x(t_n)) f_X(x) dx \quad (1.26)$$

When one ignores one stochastic function ξ_n one obtains the **reduced joint probability density**

$$\int P_n(\xi_1, t_1, \dots, \xi_n, t_n) d\xi_n = P_{n-1}(\xi_1, t_1, \dots, \xi_{n-1}, t_{n-1}) \quad (1.27)$$

The correlation between values of Ξ at different times is measured by the time-dependent moments

$$\langle \xi_1(t_1) \xi_2(t_2) \cdots \xi_n(t_n) \rangle = \int \cdots \int d\xi_1 d\xi_2 \cdots d\xi_n \xi_1 \xi_2 \cdots \xi_n P_n(\xi_1, t_1, \dots, \xi_n, t_n) \quad (1.28)$$

For **stationary processes**

$$P_n(\xi_1, t_1, \dots, \xi_n, t_n) = P_n(\xi_1, t_1 + \tau, \dots, \xi_n, t_n + \tau) \quad (1.29)$$

for all n, t_j and τ .

The conditional probability density $P_{1|1}(\xi_2, t_2 | \xi_1, t_1)$ for Ξ to take on values ξ_2 at t_2 given that it took on values ξ_1 at t_1 is defined by the joint probability

density

$$P_{1|1}(\xi_2, t_2 | \xi_1, t_1) P_1(\xi_1, t_1) = P_2(\xi_1, t_1, \xi_2, t_2) \tag{1.30}$$

In physics language this is often referred to as the **transition probability** between state 1 and state 2. Generalizing this, the conditional probability density

$$P_{m|k}(\xi_{k+1}, t_{k+1}, \dots, \xi_{k+m}, t_{k+m} | \xi_1, t_1, \dots, \xi_k, t_k) \tag{1.31}$$

for Ξ to take on the value ξ_{k+1} at $t_{k+1} \dots \xi_{k+m}$ at t_{k+m} given that it took on the value ξ_1 at t_1, \dots, ξ_k at t_k is defined by

$$P_{m|k}(\xi_{k+1}, t_{k+1}, \dots, \xi_{k+m}, t_{k+m} | \xi_1, t_1, \dots, \xi_k, t_k) \equiv \frac{P_{k+m}(\xi_1, t_1, \dots, \xi_{k+m}, t_{k+m})}{P_k(\xi_1, t_1, \dots, \xi_k, t_k)} \tag{1.32}$$

where P_{k+m} is the joint probability density.

1.6.2 Markov processes

A **Markov process** is a stochastic process where the random variable has memory only of its immediate past, i.e.

$$P_{1|n-1}(\xi_n, t_n | \xi_1, t_1, \dots, \xi_{n-1}, t_{n-1}) = P_{1|1}(\xi_n, t_n | \xi_{n-1}, t_{n-1}) \tag{1.33}$$

A Markov process is entirely determined by $P_1(\xi_1, t_1)$ and $P_{1|1}(\xi_2, t_2 | \xi_1, t_1)$. It is easy to show by using the Bayes rule that for Markov processes

$$P_{1|1}(\xi_3, t_3 | \xi_1, t_1) = \int d\xi_2 P_{1|1}(\xi_3, t_3 | \xi_2, t_2) P_{1|1}(\xi_2, t_2 | \xi_1, t_1) \tag{1.34}$$

i.e. the two steps are statistically independent. This is the **Chapman–Komogorov (CK)** or **Einstein–Smolochousky (ES)** equation. For stationary Markov processes, if we define

$$P_{1|1}(\xi_2, t_2 | \xi_1, t_1) \equiv P_\tau(\xi_2 | \xi_1) \tag{1.35}$$

since they depend only on $\tau = t_2 - t_1$, the CK or ES equation can be written schematically as

$$P_{\tau'+\tau} = P_{\tau'} P_\tau \tag{1.36}$$

in the sense of integral kernels.⁸ We now derive a differential form of the CK equation which is known as the (Markovian) Pauli master equation. Consider a small increment τ in time from t_1 and expand $P_{1|1}(\xi_2, t_1 + \tau | \xi_1, t_1)$ in a Taylor

⁸ In probability theory understandably the symbol P is used profusely. Here a single subscript τ denotes the conditional probability density $P_{1|1}$ in a stationary process. Notice this equation which we see quite commonly in physics actually presupposes the Markovian property. When it involves probability concepts, as in quantum mechanics, interpreting physics equations in the stochastic process sense may reveal a deeper layer of meaning for these common objects.

series around t_1 making sure that the normalization condition is preserved to all orders in τ ; we have

$$P_{1|1}(\xi_2, t_1 + \tau | \xi_1, t_1) = P_{1|1}(\xi_2, t_1 | \xi_1, t_1) + \tau \partial P_{1|1} / \partial \tau + \dots \tag{1.37}$$

Extract the singular part from $\partial P_{1|1} / \partial \tau$

$$\frac{\partial P_{1|1}}{\partial \tau} \equiv -a_0(\xi_1) \delta(\xi_2 - \xi_1) + W(\xi_2 | \xi_1) \tag{1.38}$$

We say $W(\xi_2 | \xi_1)$ is the **transition probability per unit time**. In physics language this is called the transition rate. $a_0(\xi_1)$ is determined by the condition that the normalization condition $\int P_{1|1}(\xi_2, t_2 | \xi_1, t_1) d\xi_2 = 1$ is satisfied to all orders of τ . To first order in τ , the condition yields

$$a_0(\xi_1) = \int W(\xi_2 | \xi_1) d\xi_2 \tag{1.39}$$

Using this we have

$$P_\tau(\xi_2 | \xi_1) = (1 - a_0(\xi_1)\tau) \delta(\xi_2 - \xi_1) + \tau W(\xi_2 | \xi_1) \tag{1.40}$$

Writing down a copy of this equation for $P_{\tau'}(\xi_3 | \xi_2)$

$$P_{\tau'}(\xi_3 | \xi_2) = (1 - a_0(\xi_2)\tau') \delta(\xi_3 - \xi_2) + \tau' W(\xi_3 | \xi_2) \tag{1.41}$$

and putting them back into the CK equation we obtain

$$\begin{aligned} P_{\tau+\tau'}(\xi_3 | \xi_1) &= \int (1 - a_0(\xi_2)\tau') \delta(\xi_3 - \xi_2) P_\tau(\xi_2 | \xi_1) d\xi_2 \\ &\quad + \int \tau' W(\xi_3 | \xi_2) P_\tau(\xi_2 | \xi_1) d\xi_2 \end{aligned} \tag{1.42}$$

Performing the integral in the first term, we obtain upon dividing by τ' on both sides and letting $\tau' \rightarrow 0$

$$\frac{\partial P_\tau(\xi_3 | \xi_1)}{\partial \tau} = \int d\xi_2 [-W(\xi_2 | \xi_3) P_\tau(\xi_3 | \xi_1) + W(\xi_3 | \xi_2) P_\tau(\xi_2 | \xi_1)] \tag{1.43}$$

where we have used the expression for a_0 above. This is the CK equation for stationary Markov process which include the familiar gain-loss, birth-death processes.

To cast this in a more familiar form we can eliminate ξ_1 by introducing the two conditional probability densities

$$P_\tau(\xi_3 | \xi_1) = P_1(\xi_3, t) \rightarrow P_n(t) \tag{1.44}$$

$$P_\tau(\xi_2 | \xi_1) = P_1(\xi_2, t) \rightarrow P_{n'}(t) \tag{1.45}$$

The right arrow indicates transforming to a notation for processes via discrete variables Ξ , as in quantum states. We get the familiar Pauli master equation

$$\frac{dP_n}{dt} = \sum_{n'} [-W_{n \rightarrow n'} P_n(t) + W_{n' \rightarrow n} P_{n'}(t)] \tag{1.46}$$

In conventional (less rigorous) physics language we call $P_n(t)$ the probability to find the system in state n , and $W_{n \rightarrow n'} P_n(t)$ the transition probability from state n to state n' in time t . Thus the first term measures the “loss” of system in state n (depletion), and the second term its “gain” (increase).⁹

1.6.3 Kramers–Moyal and Fokker–Planck equations

For linear systems and in the limit where the jumps in a Markov process are small, this equation takes a special form known as the Fokker–Planck equation. Define $\eta = \xi - \xi'$ as the jump size. The transition probability $W(\xi|\xi') = W(\xi'; \eta)$ is assumed to vary slowly with ξ, ξ' , and is a sharply peaked function of η . From the CK equation,

$$\begin{aligned} \frac{\partial P_1(\xi, t)}{\partial t} &= \int d\xi' [W(\xi|\xi') P_1(\xi', t) - W(\xi'|\xi) P_1(\xi, t)] \\ &= \int d\eta [W(\xi - \eta; \eta) P_1(\xi - \eta, t) - P_1(\xi, t) \int d\eta W(\xi - \eta; -\eta)] \end{aligned} \quad (1.47)$$

Taylor expanding P_1 around ξ in the integrand of the first term on the right-hand side, i.e.

$$P_1(\xi - \eta, t) = P_1(\xi) - \eta \frac{\partial P_1}{\partial \eta} + \frac{\eta^2}{2} \frac{\partial^2 P_1}{\partial \eta^2} + \dots \quad (1.48)$$

we have

$$\frac{\partial P_1(\xi, t)}{\partial t} = \sum_{\nu=1}^{\infty} \frac{(-1)^\nu}{\nu!} \frac{\partial^\nu}{\partial \xi^\nu} [a_\nu(\xi) P_1(\xi, t)] \quad (1.49)$$

where

$$a_\nu(\xi) = \int_{-\infty}^{\infty} \eta^\nu W(\xi; \eta) d\eta \quad (1.50)$$

This is called the Kramers–Moyal expansion of the Markovian master equation. Keeping only the first two terms and dropping the subscript 1 on P_1 (to convert

⁹ The Pauli equation could be the first instance we learn about the master equation, usually in the context of quantum mechanics (e.g. Chapter 15 of Reif [Rei67]), but it is not tied to any quantum notion whatsoever. (The only relevant concept from quantum physics is the discrete state, but we know there is a corresponding version of the CK equation for continuous variables.) To begin with, it deals with probabilities, not amplitudes, so there is no phase information, and thus is useless in dealing with issues like quantum decoherence, which probes into how the quantum phase information gets lost as a system’s classical behavior emerges. More importantly it describes only Markovian stationary process – we will see that it is far from the most general conditions. For example these are the conditions behind the Fermi Golden rule, or the Wigner-Weisskopf lineshape, which are built upon time-dependent perturbation theory. What this tells us is that it is always helpful to ask a few questions about the tacit assumptions behind any physical law, no matter how familiar they appear.

to physics notation) yields the Fokker–Planck (FP) equation

$$\frac{\partial P(\xi, t)}{\partial t} = -\frac{\partial}{\partial \xi}[a_1(\xi)P(\xi, t)] + \frac{1}{2}\frac{\partial^2}{\partial \xi^2}[a_2(\xi)P(\xi, t)] \quad (1.51)$$

For small changes in time we can write the coefficients a_1, a_2 as follows:

$$a_1(\xi) = \int \eta W(\xi, \eta) d\eta \simeq \frac{\langle \Delta \xi \rangle}{\Delta t} = \langle v_\xi \rangle \quad (1.52)$$

$$a_2(\xi) = \int \eta^2 W(\xi, \eta) d\eta \simeq \frac{\langle (\Delta \xi)^2 \rangle}{\Delta t} \quad (1.53)$$

If there is no external force $a_1 = 0$, the FP equation is in the form of a diffusion equation

$$\frac{\partial P(\xi, t)}{\partial t} = D_\xi \frac{\partial^2 P(\xi, t)}{\partial \xi^2} \quad (1.54)$$

with diffusion coefficient

$$D_\xi = \frac{a_2}{2} = \frac{\langle (\Delta \xi)^2 \rangle}{2\Delta t} \quad (1.55)$$

This is known as the first Einstein relation.

In Chapter 2 we shall use intuitive physical reasoning to give a derivation of the Boltzmann and Langevin equation, and their quantum version in Chapter 3. These equations, together with the general (not just the Pauli) master equation, will be the starting point for our expedition into nonequilibrium quantum field processes.