A Kinetic View of Statistical Physics

Aimed at graduate students, this book explores some of the core phenomena in non-equilibrium statistical physics. It focuses on the development and application of theoretical methods to help students develop their problem-solving skills.

The book begins with microscopic transport processes: diffusion, collision-driven phenomena, and exclusion. It then presents the kinetics of aggregation, fragmentation, and adsorption, where basic phenomenology and solution techniques are emphasized. The following chapters cover kinetic spin systems, from both a discrete and a continuum perspective, the role of disorder in non-equilibrium processes, hysteresis from the non-equilibrium perspective, the kinetics of chemical reactions, and the properties of complex networks. The book contains more than 200 exercises to test students' understanding of the subject. A link to a website hosted by the authors, containing an up-to-date list of errata and solutions to some of the exercises, can be found at www.cambridge.org/9780521851039.

Pavel L. Krapivsky is Research Associate Professor of Physics at Boston University. His current research interests are in strongly interacting many-particle systems and their applications to kinetic spin systems, networks, and biological phenomena.

Sidney Redner is a Professor of Physics at Boston University. His current research interests are in non-equilibrium statistical physics, and its applications to reactions, networks, social systems, biological phenomena, and first-passage processes.

Eli Ben–Naim is a member of the Theoretical Division and an affiliate of the Center for Nonlinear Studies at Los Alamos National Laboratory. He conducts research in statistical, nonlinear, and soft condensed-matter physics, including the collective dynamics of interacting particle and granular systems.

A Kinetic View of Statistical Physics

Pavel L. Krapivsky, Sidney Redner, and Eli Ben-Naim

Contents

Preface

Statistical physics is an unusual branch of science. It is not defined by a specific subject *per se*, but rather by ideas and tools that work for an incredibly wide range of problems. Statistical physics is concerned with interacting systems that consist of a huge number of building blocks – particles, spins, agents, etc. The local interactions between these elements lead to emergent behaviors that can often be simple and clean, while the corresponding few-particle systems can exhibit bewildering properties that defy classification. From a statistical perspective, the large size of a system often plays an advantageous, not deleterious, role in leading to simple collective properties.

While the tools of equilibrium statistical physics are well-developed, the statistical description of systems that are out of equilibrium is less mature. In spite of more than a century of effort to develop a formalism for non-equilibrium phenomena, there still do not exist analogs of the canonical Boltzmann factor or the partition function of equilibrium statistical physics. Moreover, non-equilibrium statistical physics has traditionally dealt with small deviations from equilibrium. Our focus is on systems far from equilibrium, where conceptually simple and explicit results can be derived for their dynamical evolution.

Non-equilibrium statistical physics is perhaps best appreciated by presenting wide-ranging and appealing examples, and by developing an array of techniques to solve these systems. We have attempted to make our treatment self-contained, so that an interested reader can follow the text with a minimum of unresolved methodological mysteries or hidden calculational pitfalls. Our main emphasis is on exact analytical tools, but we also develop heuristic and scaling methods where appropriate. Our target audience is graduate students beyond their first year who have taken a graduate course in equilibrium statistical physics and have had a reasonable exposure to mathematical techniques. We also hope that this book will be accessible to students and researchers in computer science, probability theory and applied mathematics, quantitative biological sciences, and engineering, because a wide variety of phenomena in these fields also involve the time evolution of systems with many degrees of freedom.

We begin with a few "aperitifs" – an abbreviated account of some basic problems along with some hints at methods of solution. The next three chapters comprise the major theme of transport processes. Chapter 2 introduces random walks and diffusion phenomena, mechanisms that underlie much of non-equilibrium statistical physics. Next, we discuss collision-driven phenomena in Chapter 3. We depart from the tradition of entirely focusing on the Boltzmann equation and its application to hydrodynamics. Instead, we emphasize pedagogically illuminating and tractable examples, such as the Lorentz gas and Maxwell models. In Chapter 4, we give a brief overview of exclusion processes and the profound consequences that exclusion has on transport and the spatial distribution of particles.

The next three chapters discuss the kinetics of aggregation, fragmentation, and adsorption. The classic aggregation process – in which two clusters irreversibly merge to form a larger cluster – serves as a rich playground to illustrate exact solution methods and the emergence of scaling in cluster size distributions. Many of these technical lessons will be applied throughout this book. Our presentation of the complementary process of fragmentation follows a similar logical development. We then treat the irreversible adsorption of extended objects onto a surface. Here a kinetic approach provides a remarkably easy way to solve the seemingly difficult geometric problem of determining the final coverage of the surface.

Chapters 8 and 9 are devoted to non-equilibrium spin systems. We first focus on kinetic Ising models because of their simplicity and their broad applicability to dynamic phenomena associated with phase transitions. The following chapter on coarsening develops a mesoscopic picture, in which the elemental degrees of freedom are droplets and interfaces, rather than the atomistic spins of the kinetic Ising model. These two viewpoints are complementary and each provides valuable insights. Chapter 10 gives a glimpse into the role of disorder for three specific examples of non-equilibrium processes. The next chapter exploits the insights gained from studying spin systems and disorder to treat the phenomenon of hysteresis.

Chapters 12 and 13 are devoted to population dynamics and the kinetics of chemical reactions. The first of these two chapters highlights the role of discreteness. This feature can lead to time evolution that is much different from that predicted by the deterministic rate equations. The following chapter focuses on the essential role of spatial fluctuations and dimension-dependent effects on reaction kinetics. We close with a presentation of the master equation approach to understand the basic properties of complex networks. As in the case of adsorption, the kinetic viewpoint leads to a powerful and intuitive way to determine many geometrical properties of networks.

We conclude each chapter with a short "Notes" section that provides a guide to additional reading. We direct the reader to books and review articles whenever possible. By this emphasis, we do not mean to slight original literature, but most relevant information can be found within these more comprehensive references. However, we do cite original sources when such an exposition is particularly useful pedagogically or when a particular subject has not yet been reviewed.

Our choice of topics has been guided by the desire to provide key ideas and core techniques that will help turn students of non-equilibrium statistical physics into practitioners. Owing to space limitations as well as our own personal biases and lack of knowledge, many important topics have been omitted. We hope that a student who successfully studies from this book will then be ready to competently assimilate many other topics in non-equilibrium statistical physics by self-study.

Although our coverage of topics is incomplete, the contained material is still too ambitious for a one-semester course. For such a course, we recommend most of Chapter 2 (random walks/diffusion), the first three sections of Chapter 3 (collisions), the first four sections of Chapter 5 (aggregation), sections 7.1 and 7.4 in Chapter 7, most of Chapters 8 and 9 (spin systems and coarsening), the first two sections of Chapter 12 (population dynamics), the first three sections of Chapter 13 (diffusive reactions), and Chapter 14 (complex networks). Students are encouraged to solve the problems; this is perhaps the most effective way to

learn the material. In our experience, several sections and chapters are also well-suited for stand-alone mini-courses and summer schools.

We owe a great debt of gratitude to numerous collaborators, colleagues, and students who have helped shape our thinking and who have also provided advice in the preparation of this book. Each of us has benefited from insights learned from long-term collaborators, and some of their insights have percolated their way into this book. We do not mention them by name because they are too numerous and we are sure to miss some. Nevertheless, we are truly grateful to them, and we are lucky to count many of these colleagues and co-authors among our friends.

Conventions

One of the pleasures of working in statistical physics is that there is almost no need to specify the units; it is sufficient to remember that (in classical non-relativistic physics) there are only three independent units of mass, length, and time. This is a welcome return to normality, especially after going through electrodynamics and magnetism, a beautiful subject that has been muddled by the SI system of units. In statistical physics there is one such intruder from the SI system, Boltzmann's constant k, and many physicists and even mathematicians proudly display this worthless conversion factor. Boltzmann's constant is a lesser evil than artifices such as the permittivity or the permeability of the vacuum, but keeping k is akin to carrying the conversion factor between miles and kilometers. Throughout this book the Boltzmann constant is always one, $k = 1$; that is, temperature is measured in energy units.

Non-equilibrium statistical physics has a substantial overlap with applied mathematics and we follow valuable lessons from the latter, e.g. we try to work with equations in dimensionless form. Whenever we use dimensionful equations, we try to employ dimensional analysis which often leads to spectacular advances with minimal effort. Dimensional analysis originated in hydrodynamics, the subject that is also on the cross-field between non-equilibrium statistical physics and applied mathematics.

For the sake of notational simplicity, we commit some abuses in notation. For example, we use the same letter for a function and its Fourier transform, say $f(x)$ and $f(k)$, respectively. By context, there is rarely a risk of confusion and this convention allows us to avoid cluttered notation. Similarly, we write the Laplace transform of a function $f(x)$ as $f(s)$; for the Mellin transform of $f(x)$ we also write $f(s)$. Since the Laplace and Mellin transforms never appear together, the redundancy of notation should not cause any confusion.

Some technical conventions: throughout this book an overdot always means differentiation with respect to time. The meaning of the relational symbols $\equiv$, $\cong$, $\simeq$, $\sim$, $\propto$, $\approx$, etc., is usually fuzzy in the minds of every researcher, and we can attest that it has caused us headaches throughout the preparation of this book. To keep mathematical relations simple and minimal, we tried to avoid the symbols $\cong$, $\propto$, and $\approx$ because they are not far from the symbols $\equiv$, $\simeq$, and $\sim$. Specifically:

$\cong$ has a meaning intermediate between $\equiv$ and $\simeq$, so we use one of the extremes;
$\propto$ has a meaning that resembles $\sim$, so we use $\sim$;
$\approx$ often gives the approximate numerical value of a constant; since nowadays we can easily extract as many digits as we like, there is minimal need for this symbol.

Thus we primarily use the symbols that are defined as follows:

$\equiv$ indicates "defined as";

$\simeq$ indicates "asymptotically equal to"; for example, $\sin 2x \simeq 2x$ when $x \to 0$;

$\sim$ means "proportional" (say $\sin 2x \sim x$); more generally, $\sim$ means "of the order of"; of course, we prefer the standard equal sign whenever we are writing an exact solution;

$\doteq$ means equality to as many digits as we like, but we only quote the first few; for example, $\sqrt{2} \doteq 1.41421356$.

1 Aperitifs

Non-equilibrium statistical physics describes the time evolution of many-particle systems. The individual particles are elemental interacting entities which, in some situations, can change in the process of interaction. In the most interesting cases, interactions between particles are strong and hence a deterministic description of even a few-particle system is beyond the reach of any exact theoretical approach. On the other hand, many-particle systems often admit an analytical statistical description when their number becomes large. In that sense they are *simpler* than few-particle systems. This feature has several different names – the law of large numbers, ergodicity, etc. – and it is one of the reasons for the spectacular successes of statistical physics and probability theory.

Non-equilibrium statistical physics is also quite different from other branches of physics, such as the "fundamental" fields of electrodynamics, gravity, and high-energy physics that involve a reductionist description of few-particle systems, as well as applied fields, such as hydrodynamics and elasticity, that are primarily concerned with the consequences of fundamental governing equations. Some of the key and distinguishing features of non-equilibrium statistical physics include the following:

- there are no basic equations (like Maxwell equations in electrodynamics or Navier–Stokes equations in hydrodynamics) from which the rest follows;
- it is intermediate between fundamental and applied physics;
- common underlying techniques and concepts exist in spite of the wide diversity of the field;
- it naturally leads to the creation of methods that are useful in applications outside of physics (for example the Monte Carlo method and simulated annealing).

Our guiding philosophy is that, in the absence of underlying principles or governing equations, non-equilibrium statistical physics should be oriented toward explicit and illustrative examples rather than attempting to develop a theoretical formalism that is still incomplete.

Let's start by looking briefly at the random walk to illustrate a few key ideas and to introduce several useful analysis tools that can be applied to more general problems.

1.1 Diffusion

For the symmetric diffusion on a line, the probability density

$$\text{Prob}\,[\text{particle} \in (x, x + dx)] \equiv P(x, t)\,dx \tag{1.1}$$

satisfies the diffusion equation

$$\frac{\partial P}{\partial t} = D \frac{\partial^2 P}{\partial x^2}. \tag{1.2}$$

As we discuss soon, this equation describes the continuum limit of an unbiased random walk. The diffusion equation must be supplemented by an initial condition that we take to be $P(x, 0) = \delta(x)$, corresponding to a walk that starts at the origin.

Dimensional analysis

Let's pretend that we don't know how to solve (1.2) and try to understand the behavior of the walker without explicit solution. What is the mean displacement $\langle x \rangle$? There is no bias, so clearly

$$\langle x \rangle \equiv \int_{-\infty}^{\infty} x P(x, t) \, dx = 0 \,.$$

The next moment, the mean-square displacement,

$$\langle x^2 \rangle \equiv \int_{-\infty}^{\infty} x^2 P(x, t) \, dx$$

is non-trivial. Obviously, it should depend on the diffusion coefficient D and time t. We now apply dimensional analysis to determine these dependences. If L denotes the unit of length and T denotes the time unit, then from (1.2) the dimensions of $\langle x^2 \rangle, D$, and t are

$$[\langle x^2 \rangle] = L^2 \,, \quad [D] = L^2/T \,, \quad [t] = T \,.$$

The only quantities with units of length squared that can be formed from these parameters are the mean-square displacement itself and the product Dt. Hence

$$\langle x^2 \rangle = C \times Dt. \tag{1.3}$$

Equation (1.3) is one of the central results in non-equilibrium statistical physics, and we derived it using just dimensional analysis! To determine the numerical constant $C = 2$ in (1.3) one must work a bit harder (e.g. by solving (1.2), or by multiplying Eq. (1.2) by x^2 and integrating over the spatial coordinate to give $d\langle x^2 \rangle/dt = 2D$).

This seemingly trivial reasoning works for much more complex problems, such as estimating the age of the universe from its density or the energy of an atomic blast from the time dependence of the blast wave radius. Because of its simplicity and power, dimensional analysis should be used as the first resort for investigating real problems.

Scaling

Let's now apply dimensional analysis to the probability density $P(x, t | D)$; here D is explicitly displayed to remind us that the density does depend on the diffusion coefficient. Since

$[P] = L^{-1}$, the quantity $\sqrt{Dt}\, P(x, t \,|\, D)$ is dimensionless, so it must depend on dimensionless quantities only. From variables x, t, D we can form a single dimensionless quantity $x/\sqrt{Dt}$. Therefore the most general dependence of the density on the basic variables that is allowed by dimensional analysis is

$$P(x, t) = \frac{1}{\sqrt{Dt}} \, \mathcal{P}(\xi), \qquad \xi = \frac{x}{\sqrt{Dt}} \, . \tag{1.4}$$

The density depends on a single *scaling variable* rather than on two basic variables x and t. This remarkable feature greatly simplifies analysis of the typical partial differential equations that describe non-equilibrium systems. Equation (1.4) is often referred to as the *scaling ansatz*. Finding the right scaling ansatz for a physical problem often represents a large step toward a solution. For the diffusion equation (1.2), substituting in the ansatz (1.4) reduces this partial differential equation to the ordinary differential equation

$$2\mathcal{P}'' + \xi \mathcal{P}' + \mathcal{P} = 0 \, .$$

Integrating twice and invoking both symmetry ($\mathcal{P}'(0) = 0$) and normalization, we obtain $\mathcal{P} = (4\pi)^{-1/2} \, e^{-\xi^2/4}$, and finally the Gaussian probability distribution

$$P(x, t) = \frac{1}{\sqrt{4\pi Dt}} \, \exp\left[-\frac{x^2}{4Dt} \right] \, . \tag{1.5}$$

In this example, the scaling form was rigorously derived from dimensional reasoning. In more complicated situations, arguments in favor of scaling are less rigorous, and scaling is usually achieved only in some asymptotic limit. The above example where scaling applies for all t is an exception; for the diffusion equation with an initial condition on a finite domain rather than a point support, scaling holds only in the limit $x, t \to \infty$, with the scaling variable ξ kept finite. Nevertheless, we shall see that, whenever it is applicable, scaling immediately advances the understanding of a problem.

Renormalization

The strategy of the renormalization group method is to understand the behavior on large "scales" – here meaning long times – by iterating the properties of random walks on smaller time scales. For the diffusion equation, we start with the identity

$$P(x, 2t) = \int_{-\infty}^{\infty} P(y, t) \, P(x - y, t) \, dy, \tag{1.6}$$

which expresses the probability that the walk is at position x at time $2t$ in terms of the propagation from time 0 to t and from time t to $2t$. Mathematically, the probability distribution after time $2t$ is given by the convolution of probability distributions to reach the intermediate time t and the probability distribution to propagate from time t to $2t$. Equation (1.6) is known as a Chapman–Kolmogorov equation and it reflects the memoryless Markov nature of the random walk. (Equation (1.6) is also the basis for the path integral treatment of diffusion processes but we will not delve into this subject here.)

The convolution form of Eq. (1.6) calls out for applying the Fourier transform,

$$P(k,t) = \int_{-\infty}^{\infty} e^{ikx} P(x,t)\, dx, \qquad (1.7)$$

that recasts (1.6) into the algebraic relation $P(k, 2t) = [P(k,t)]^2$. The scaling form (1.4) shows that $P(k,t) = \mathcal{P}(\kappa)$ with $\kappa = k\sqrt{Dt}$, so the renormalization group equation is

$$\mathcal{P}(\sqrt{2}\,\kappa) = [\mathcal{P}(\kappa)]^2.$$

Taking logarithms and defining $z \equiv \kappa^2$, $Q(z) \equiv \ln \mathcal{P}(\kappa)$, we arrive at $Q(2z) = 2Q(z)$, whose solution is $Q(z) = -Cz$, or $P(k,t) = e^{-2k^2 Dt}$. (The constant $C = 2$ may be found, e.g. by expanding (1.7) for small k, $P(k,t) = 1 - k^2 \langle x^2 \rangle$, and recalling that $\langle x^2 \rangle = 2Dt$.) Performing the inverse Fourier transform we recover (1.5). Thus the Gaussian probability distribution represents an exact solution to a renormalization group equation. This derivation shows that the renormalization group is ultimately related to scaling.

Master equation approach

The *master equation* provides a way to describe the evolution of many non-equilibrium systems. Such systems may be described statistically by the probability that its state variables have specified values. The rate at which these state variables change in time is accounted for by the master equation. Let us illustrate this equation with the simple example of a symmetric random walk on a one-dimensional lattice. The walk is described by the probability $P_n(t)$ that the walk is at site n at time t. This probability evolves as

$$\frac{\partial P_n}{\partial t} = P_{n-1} + P_{n+1} - 2P_n. \qquad (1.8)$$

The first two terms on the right account for the increase in P_n because of a hop from $n-1$ to n or because of a hop from $n+1$ to n, respectively. Similarly, the last term accounts for the decrease of P_n because of hopping from n to $n \pm 1$.

As we will discuss throughout this book, random walks underlie many aspects of non-equilibrium statistical physics. Chapter 2 will be devoted to understanding random walks and related diffusion processes. We will also present, in Chapters 2 and 10, physically motivated generalizations, including: (i) global bias in the hopping rates, (ii) long-range hopping, and (iii) hopping rates that are random functions of position. The insights gained from studying random walks will provide essential background for dealing with the governing equations that describe the evolution of non-equilibrium processes in interacting many-body systems.

1.2 Single-species annihilation/coalescence

In non-equilibrium statistical physics, we study systems that contain a macroscopic number of interacting particles. To understand collective behaviors it is useful to ignore complications resulting from finiteness, i.e. to focus on situations when the number of particles is

infinite. Perhaps the simplest interacting infinite-particle systems of this kind are *single-species annihilation*, where particles diffuse freely and annihilate instantaneously upon contact, and *single-species coalescence*, where the reactants merge upon contact. These processes have played an important role in the development of non-equilibrium statistical physics and they provide excellent illustrations of techniques that can be applied to other infinite-particle systems.

These two processes are symbolically represented by the respective reaction schemes

$$A + A \longrightarrow \emptyset, \tag{1.9a}$$

$$A + A \longrightarrow A. \tag{1.9b}$$

The density $n(t)$ of A particles for both reactions obviously decays with time. The question is: how?

Hydrodynamic description

In the hydrodynamic approach, one assumes that the reactants are perfectly mixed at all times. This means that the density at every site is the same and that every particle has the same probability to react at the next instant. In this well-mixed limit, and also assuming the continuum limit, the particle density ρ in the coalescence reaction decays with time according to the *rate equation*

$$\frac{d\rho}{dt} = -K\rho^2 . \tag{1.10}$$

Indeed, since two particles have to be at the same location for a reaction to occur, the right-hand side is proportional to the density squared. Here K is the reaction rate that describes the propensity for two diffusing particles to interact; the computation of this rate requires a detailed microscopic treatment (see Section 2.7). For the annihilation reaction, we can use essentially the same governing equation, namely $d\rho/dt = -2K\rho^2$, with now a prefactor of 2 to account for two particles being eliminated in each collision.

Equations of the form (1.10) are typical of a hydrodynamic description; the solution to this equation is

$$n(t) = \frac{\rho_0}{1 + K\rho_0 t} \sim (Kt)^{-1} , \tag{1.11}$$

where ρ_0 is the initial density. However, the true asymptotic behavior is more interesting and depends on the spatial dimension d:

$$\rho(t) \sim \begin{cases} t^{-1/2}, & d = 1, \\ t^{-1} \ln t, & d = 2, \\ t^{-1}, & d > 2. \end{cases} \tag{1.12}$$

The change at $d_c = 2$ illustrates the important notion of the critical dimension: above d_c, the rate equation description leads to asymptotically correct behavior; below d_c, this approach

is wrong; at d_c, this approach is almost correct – it typically is in error by a logarithmic correction term.

To obtain a complete theory of the reaction, one might try to write formally exact equations for correlation functions. That is, if the two-argument function $\rho(\mathbf{r}, t)$ is the microscopic density, the true dynamical equation for $\rho(t) \equiv \langle \rho(\mathbf{r}, t) \rangle$ involves the second-order correlation functions (or correlators) $\langle \rho(\mathbf{r}, t)\rho(\mathbf{r}', t) \rangle$ since two particles need to be in proximity for a reaction to occur. Similarly, the equation for the second-order correlation functions involves third-order correlators, etc. These equations are hierarchical and the only way to proceed is to impose some sort of closure scheme in which higher-order correlators are factorized in terms of lower-order correlators. In particular, the hydrodynamic equation (1.10) is recovered if we assume that second-order correlators factorize; that is, $\langle \rho(\mathbf{r}, t)\rho(\mathbf{r}', t) \rangle = \langle \rho(\mathbf{r}, t) \rangle \langle \rho(\mathbf{r}', t) \rangle = \rho(t)^2$. Thus Eq. (1.10) is the factorized version of the Boltzmann equation for the annihilation process (1.9a). Attempts to describe this reaction scheme more faithfully by higher-order correlators have not been fruitful. Thus the revered kinetic theory approach is essentially useless for the innocent-looking annihilation process $A + A \to \emptyset$! Because of this failure of approaches that are based on traditional calculations, alternative methods are needed.

Dimensional analysis

Let's determine the dependence of the rate K on fundamental parameters of the reaction, i.e. on the diffusion coefficient D of the reactants and radius R of each particle. From Eq. (1.10), the dimension of the reaction rate is $[K] = L^d/T$, and hence the only possible dependence is[1]

$$K = DR^{d-2}. \tag{1.13}$$

Using (1.13) in (1.10) and solving this equation yields

$$\rho(t) \sim \frac{1}{R^{d-2}Dt}. \tag{1.14}$$

We anticipate that the density ought to decay more quickly when the radius of the particles is increased. According to (1.14), this anticipation is true only when $d > 2$. Thus the rate equation could be correct in this regime. For $d = 2$, the decay is independent of the size of particles – already a surprise. For $d < 2$, we obtain the obviously wrong result that the density decays more slowly if particles are larger.

The density is actually *independent* of R for $d < 2$. This feature is self-evident for $d = 1$ because all that matters is the spacing between particles. If we now seek, on dimensional grounds, the density in the R-independent form $\rho = \rho(D, t)$, we find that the only possibility is $\rho \sim (Dt)^{-d/2}$, in agreement with the prediction of (1.12) in one dimension. In the context

---

[1] Here we omit a numerical factor of the order of one that cannot be determined within the realm of dimensional analysis; in the future, we shall often ignore such factors without explicit warning. Note that the reaction rate is not proportional to the cross-sectional area, R^{d-1}, but rather to R^{d-2}; this feature stems from the vagaries of diffusive motion; see also Section 2.7.

of the reaction rate, this slow decay is equivalent to a reaction rate that decreases with time. We will return to this point in the next chapter.

Heuristic arguments

Dimensional analysis often gives correct dependences but does not really explain why these behaviors are correct. For the annihilation process (1.9a), we can understand the one-dimensional asymptotic, $\rho \sim (Dt)^{-1/2}$, in a physical way by using a basic feature (1.3) of random walks: in a time interval t, each particle explores the region $\ell \sim \sqrt{Dt}$. Therefore the typical separation between surviving particles is of the order of ℓ, from which $\rho \sim \ell^{-1} \sim (Dt)^{-1/2}$ follows.

Guided by this understanding, let's try to understand (1.12) in all spatial dimensions. It is helpful to consider a slightly modified situation in which particles undergo random walks on a lattice in d dimensions (the lattice spacing plays the role of the particle radius). What is the average number of *distinct* sites $\mathcal{N}$ visited by a random walker after N steps? This question has a well-known and beautiful answer:

$$\mathcal{N} \sim \begin{cases} N^{1/2}, & d = 1, \\ N/\ln N, & d = 2, \\ N, & d > 2. \end{cases} \tag{1.15}$$
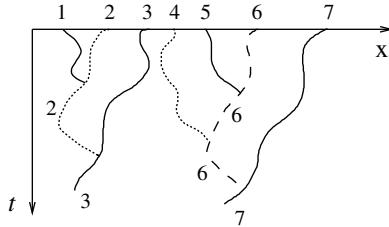
With a little contemplation, one should be convinced that the density in single-species annihilation scales as the inverse of the average number of sites visited by a random walker; if there is more that one particle in the visited region, it should have been annihilated previously. Thus (1.15) is equivalent to (1.12).

Exact solution in one dimension

Diffusion-controlled annihilation admits an exact solution in one dimension. This is an exceptional feature – most interacting infinite-particle systems cannot be solved, even in one dimension. Moreover, even in solvable cases, we can usually compute only a limited number of quantities. For one-dimensional annihilation, for example, while the density is known exactly, the distribution of distances ℓ between adjacent particles $P(\ell, t)$ is unknown, even in the simpler scaling limit $\ell, t \rightarrow \infty$, with $\xi = \ell/\sqrt{Dt}$ finite, where $P(\ell, t)$ can be written as $(Dt)^{-1} \mathcal{P}(\xi)$, with $\mathcal{P}$ as yet unknown.[2]

Exact results for diffusion-controlled annihilation will be presented in Chapters 8 and 13, where we develop the necessary technical tools. For now, let's consider diffusion-controlled coalescence, $A + A \longrightarrow A$, that *is* readily soluble in one dimension because it can be reduced to a two-particle problem. To compute the density, it is convenient to define particle labels so that in each collision the left particle disappears and the right particle survives (Fig. 1.1). Then to compute the survival probability of a test particle we may ignore all particles to the left. This reduction of the original two-sided problem to a one-sided problem considerably

---

[2] The $(Dt)^{-1}$ prefactor ensures that the "conservation law" $\int_0^\infty \ell P(\ell, t) \, d\ell = 1$ can be written in the manifestly time-independent form $\int_0^\infty \xi \mathcal{P}(\xi) \, d\xi = 1$.

Space-time trajectories of diffusing particle in one dimension, showing the propagation of particle labels in diffusion-controlled coalescence, $A + A \rightarrow A$.

simplifies the analysis because only the closest particle to the right of the test particle is relevant. While the right neighbor can merge with other particles further to the right, these reactions never affect the fate of the test particle. Thus the system reduces to a soluble two-particle problem.

The distance between the test particle and its right neighbor undergoes diffusion with diffusivity $2D$ because the spacing diffuses at twice the rate of each particle. Consequently, the probability density $p(\ell, t)$ that the test particle is separated by distance ℓ from its right neighbor satisfies the diffusion equation subject to the absorbing boundary condition:

$$\frac{\partial p}{\partial t} = 2D \frac{\partial^2 p}{\partial \ell^2}, \qquad \text{with} \quad p(0, t) = 0. \tag{1.16}$$

The solution to (1.16) for an arbitrary initial condition $p(y, t=0)$ is

$$p(\ell, t) = \frac{1}{\sqrt{8\pi Dt}} \int_0^\infty p(y, t=0) \, [e^{-(\ell-y)^2/8Dt} - e^{-(\ell+y)^2/8Dt}] \, dy$$

$$= \frac{1}{\sqrt{2\pi Dt}} \exp\left(-\frac{\ell^2}{8Dt}\right) \int_0^\infty p(y, t=0) \, \exp\left(-\frac{y^2}{8Dt}\right) \sinh\left(\frac{\ell y}{4Dt}\right) dy. \tag{1.17}$$

In the first line, the solution is expressed as the superposition of a Gaussian and an image anti-Gaussian that automatically satisfies the absorbing boundary condition. In the long-time limit, the integral on the second line tends to

$$\frac{\ell}{4Dt} \int_0^\infty y \, p(y, t = 0) \, yy = \frac{\ell}{4Dt\rho_0} \, ,$$

where ρ_0 is the initial density. Therefore

$$p(\ell, t) \simeq \frac{\ell}{4Dt\rho_0\sqrt{2\pi Dt}} \exp\left(-\frac{\ell^2}{8Dt}\right) \, ,$$

leading to the asymptotic density decay

$$\rho(t) = \rho_0 \int_0^\infty p(\ell, t) \, d\ell \simeq (2\pi Dt)^{-1/2} \, , \tag{1.18}$$

which is independent of the initial density.

Since the interval length distribution $P(\ell, t)$ is equal to $\rho_0\, p(\ell, t)$, we see that for diffusion-controlled coalescence the scaled distribution reads

$$\mathcal{P}(\xi) = \frac{\xi}{4\sqrt{2\pi}}\, \exp\!\left(-\frac{\xi^2}{8}\right), \qquad \xi = \frac{\ell}{\sqrt{Dt}}\,.$$

1.3 Two-species annihilation

Consider two diffusing species A and B that are initially distributed at random with equal concentrations: $\rho_A(0) = \rho_B(0) = \rho_0$. When two particles of opposite species approach within the reaction radius, they immediately annihilate according to

$$A + B \longrightarrow \emptyset\,. \tag{1.19}$$

For this reaction, the density decreases as

$$\rho(t) \sim \begin{cases} t^{-d/4}, & d \le 4, \\ t^{-1}, & d > 4, \end{cases} \tag{1.20}$$

as $t \to \infty$, so the critical dimension is $d_c = 4$. This result shows that the hydrodynamic description is wrong even in the physically relevant case of three dimensions. A peculiar feature of Eq. (1.20) is the absence of logarithmic corrections for $d = d_c$. Typically, logarithmic corrections arise at the critical dimension and their absence for two-species annihilation is, as yet, unexplained.
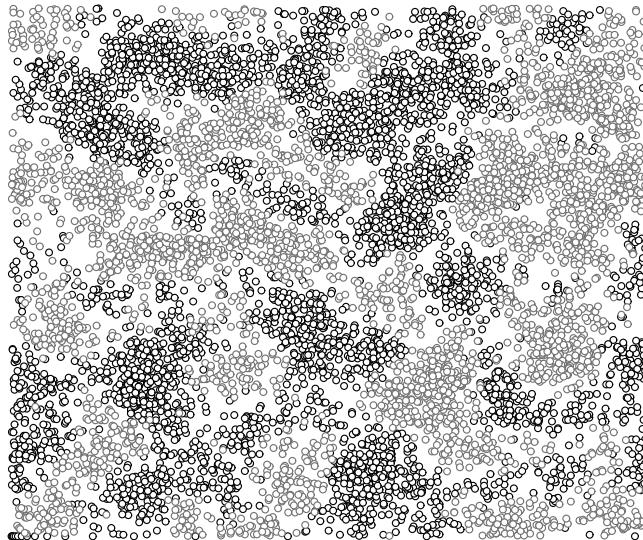
In this example, neither a hydrodynamic description (that gives $\rho \sim t^{-1}$) nor dimensional analysis can explain the decay of the density. To understand why the naive approaches fail, consider a snapshot of a two-dimensional system at some time $t \gg 1$ (Fig. 1.2). The system spontaneously organizes into a mosaic of alternating A and B domains. Because of this organization, annihilation can occur only along domain boundaries rather than throughout the system. This screening effect explains why the density is larger than in the hydrodynamic picture, where particles are taken to be well mixed.

A heuristic argument that is based on this domain picture allows us to determine the density decay in Eq. (1.20). In a spatial region of linear size ℓ, the initial number of A particles is

$$N_A = \rho_0 \ell^d \pm \sqrt{\rho_0 \ell^d}\,, \tag{1.21}$$

and similarly for B particles. Here the $\pm$ term signifies that the particle number in a finite region is a stochastic variable that typically fluctuates in a range of the order of $\sqrt{\rho_0 \ell^d}$ about the mean value $\rho_0 \ell^d$. Consequently, the typical value of the difference $N_A - N_B$ in a region of size ℓ is

$$N_A - N_B = \pm\sqrt{\rho_0 \ell^d}\,.$$

Snapshot of the particle positions in two-species annihilation in two dimensions.

This number difference is not affected by annihilation events. Therefore after the minority species in a given region is eliminated, the local density becomes $\rho \sim \sqrt{\rho_0 \ell^d}/\ell^d$. Because of the diffusive spreading (1.3), the average domain size scales as $\ell \sim \sqrt{Dt}$, and hence

$$\rho(t) \sim \sqrt{\rho_0}\,(Dt)^{-d/4}\,. \tag{1.22}$$

In contrast to the single-species annihilation where the initial density is asymptotically irrelevant, the initial density affects the asymptotic behavior of two-species annihilation.

Finally, notice that the density decay cannot be obtained by dimensional analysis alone because now there are at least two independent length scales, the domain size $\sqrt{Dt}$ and the interparticle spacing. Additional physical input, here in the form of the domain picture, is needed to obtain the time dependence of the density.

1.4 Notes

There is considerable literature on the topics discussed in this introductory chapter. Random walks and diffusion are classic subjects for which essential background and additional references can be found in [1–6]. Dimensional analysis and the concept of scaling go back to Galileo [7]. A modern perspective and the historical roots of scaling are presented in [8]. A field where dimensional analysis and scaling are especially popular is hydrodynamics. An informative exposition appears in the book by Barenblatt [9] that discusses the connection between scaling, intermediate asymptotics, and the renormalization group. The latter connections are also emphasized in Goldenfeld's book [10]. Another very useful book that emphasizes the scaling approach for polymer physics is that of de Gennes [11]. The kinetics

of reaction processes are discussed in Chapters 12 and 13, and literature suggestions in this subject will be given in those chapters. The time dependence (1.15) of the number of distinct sites visited by a random walk is a classic result and is derived, for example, in [1] and [3]. The diffusion-controlled coalescence process in one dimension is one of the few examples of an irreversible reaction that can justifiably be called "completely solvable"; references to the various exact solutions of this model and its generalizations will also be given in Chapters 12 and 13.

2 Diffusion

The random walk and its continuum limit of diffusion, or Brownian motion, lie at the heart of non-equilibrium statistical physics. The mechanism for Brownian motion[1] is the repeated collisions experienced by a microscopic particle with molecules in the surrounding environment. The historical importance of Brownian motion is that it provided a strong (indirect) confirmation of the existence of atoms and molecules. Because of their ubiquity and utility, random walks (Fig. 2.1) and diffusion are arguably the most successful models of theoretical physics, with applications far wider than merely describing the motion of microscopic particles. Here we give a brief introduction to basic results and applications that will be used throughout this book.

2.1 The probability distribution

We start by considering a discrete-time random walk on a one-dimensional lattice. At each step, the walker moves a unit distance to the right with probability p or a unit distance to the left with probability $q = 1 - p$. The probability $P_N(x)$ that the walk is at x at the Nth

(a)   (b)
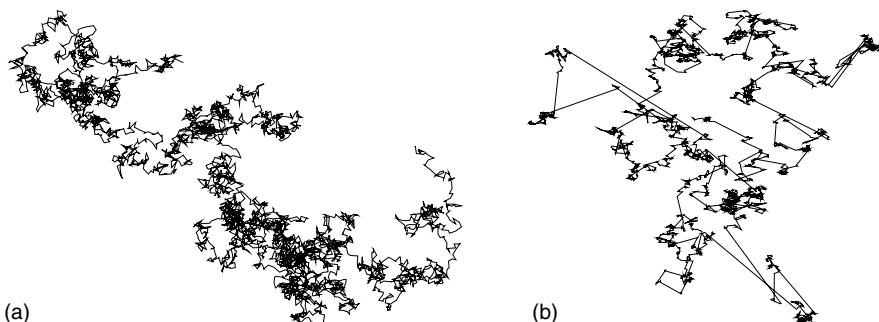
Fig. 2.1. (a) A 5000-step simple random walk in which each step is of fixed length and random direction in two dimensions. (b) A 5000-step random walk with a distribution of step lengths r that scales as $r^{-4/3}$.

[1] Brownian motion was discovered in 1785 by Jan Ingenhousz, who noticed the irregular motion of coal dust particles on the surface of alcohol. Robert Brown made his observations in 1827 in studying the movement of suspended pollen grains in water.

step obeys the recursion

$$P_N(x) = pP_{N-1}(x-1) + qP_{N-1}(x+1).$$ (2.1)

Instead of directly solving (2.1), we notice that the probability $\Pi_N(r)$ that the walk takes r steps to the right and $N-r$ to the left has the *binomial* form:

$$\Pi_N(r) = \binom{N}{r} p^r q^{N-r}.$$ (2.2)

The binomial factor $\binom{N}{r}$ counts the number of different arrangements of right and left steps that results in r total steps to the right, while the factor $p^r q^{N-r}$ gives the probability for a single walk. If the random walk starts at the origin, the total displacement will be $x = 2r - N$. Using Stirling's approximation for large N to simplify the binomial distribution we find that $P_N(x) = \Pi_N[(x+N)/2]$ becomes[2]

$$P_N(x) \rightarrow \frac{1}{\sqrt{2\pi Npq}} e^{-[x-N(p-q)]^2/2Npq}.$$ (2.3)

This Gaussian form arises universally when the mean displacement and the mean-square displacement in a single step are finite. We will return to this point in our treatment of the central limit theorem in Section 2.2.

Discrete-time random walks are important in mathematics and computer science, but in the natural sciences, continuous-time random walks usually arise. Let's study the symmetric case where hopping to the left or to the right occur with equal rates that we set to 1 without loss of generality (Fig. 2.2). Denote by $P_n(t)$ the probability for the walk to be at site n at time t. The *master equation* for this occupation probability is

$$\frac{\partial P_n}{\partial t} = P_{n+1} - 2P_n + P_{n-1}.$$ (2.4)

Namely, P_n increases due to the hopping from $n \pm 1$ to n; conversely, P_n decreases because of hopping from n to $n-1$ and to $n+1$. The solution to Eq. (2.4) is (see the highlight below, specifically Eq. (2.15))

$$P_n(t) = I_n(2t) e^{-2t},$$ (2.5)
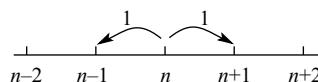
Fig. 2.2.   A continuous-time random walk that hops to the left and to the right with rate 1.

---

[2] More precisely, $P_N(x)$ is twice larger than the prediction of (2.3) when x and N have the same parity and $P_N(x) \equiv 0$ if x and N have opposite parities.

where I_n is the modified Bessel function of order n. In the long-time limit, the asymptotics of the Bessel function leads to the Gaussian distribution

$$P_n(t) \to \frac{1}{\sqrt{4\pi t}}\, e^{-n^2/4t}\,. \tag{2.6}$$

In continuous space, the governing equation (2.4) turns into the *diffusion equation*, a classical equation of physics that underlies much of this book. To make space continuous, we replace $n \to x$ and $P_n(t)$ by $P(x,t)$, the probability density $P(x,t)$ at x at time t. Next we expand (2.4) in a Taylor series to second order. These steps give the diffusion equation

$$\frac{\partial P(x,t)}{\partial t} = D\, \frac{\partial^2 P(x,t)}{\partial x^2}, \tag{2.7}$$

where the diffusion coefficient $D = 1$. For general D, we solve this diffusion equation by introducing the Fourier transform

$$P(k,t) = \int_{-\infty}^{\infty} P(x,t)\, e^{ikx}\, dx, \qquad P(x,t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} P(k,t)\, e^{-ikx}\, dk\,. \tag{2.8}$$

For completeness, we have also written the inverse Fourier transform. As we shall see throughout this chapter, the use of the Fourier transform simplifies the analysis of the equations that arise in random walks. For the present example, the Fourier transform recasts (2.7) into

$$\frac{\partial P(k,t)}{\partial t} = -Dk^2\, P(k,t),$$

with solution $P(k,t) = P(k,0)\, e^{-Dk^2t}$. Using the initial condition, $P(x,t=0) = \delta(x)$, we get $P(k,t) = e^{-Dk^2t}$. We then invert the Fourier transform and obtain the Gaussian probability distribution[3]

$$P(x,t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-Dk^2t}\, e^{-ikx}\, dk = \frac{1}{\sqrt{4\pi Dt}}\, e^{-x^2/4Dt}, \tag{2.9}$$

which is identical to the discrete-space result (2.6) when the diffusion coefficient is set to 1.

Discrete hopping equation and modified Bessel functions

Let's study a slightly more general version of Eq. (2.4):

$$\frac{\partial P_n}{\partial t} = \gamma\,(P_{n-1} + P_{n+1}) - 2P_n\,. \tag{2.10}$$

This generalization describes the evolution of spin correlation functions in the kinetic Ising model (to be discussed in Chapter 8). For simplicity, suppose that the initial condition is

---

[3] The integral $\int_{-\infty}^{\infty} e^{-ak^2-ikx}\, dk$ is performed by "completing the square" in the exponential. We write $ak^2+ikx = a(k + ix/2a)^2 - x^2/4a$, transform to $y = k + ix/2a$, and recall that $\int_{-\infty}^{\infty} e^{-ay^2}\, dy = \sqrt{\pi/a}$.

$P_n(t = 0) = \delta_{n,0}$. To solve (2.10), introduce the Fourier transform

$$P(k, t) = \sum_{n=-\infty}^{\infty} P_n(t) \, e^{ikn},\tag{2.11}$$

in terms of which the equation of motion becomes

$$\frac{dP(k, t)}{dt} = [\gamma(e^{ik} + e^{-ik}) - 2] P(k, t).\tag{2.12}$$

Since the Fourier transform of the initial condition is $P(k, t = 0) = 1$, the solution to Eq. (2.12) is

$$P(k, t) = e^{2(\gamma \cos k - 1)t}.\tag{2.13}$$

The generating function for the modified Bessel function I_n has the compact expression

$$e^{z \cos k} = \sum_{n=-\infty}^{\infty} e^{ikn} \, I_n(z).\tag{2.14}$$

By comparing this generating function with the solution (2.13) for $P(k, t)$ and using definition (2.11) of the Fourier transform, we immediately read off the solution:

$$P_n(t) = I_n(2\gamma t) \, e^{-2t}.\tag{2.15}$$

2.2 Central limit theorem

After many steps, the probability distribution for the displacement of a random walk converges to a Gaussian distribution, *independent* of the form of the single-step distribution (subject to mild restrictions to be discussed below). This universality is embodied by the *central limit theorem*. Suppose that the walk is confined to one dimension[4] and takes steps at discrete times, with the displacement of the nth step, x_n, drawn independently from a continuous distribution $p(x)$. Let $X_N = \sum_{1 \le n \le N} x_n$ be the displacement of the walk after N steps.[5] The central limit theorem asserts that, in the $N \to \infty$ limit, the probability distribution of the total displacement, $P_N(X)$, is the *universal Gaussian* function

$$P_N(X) \sim \frac{1}{\sqrt{2\pi N \sigma^2}} \, e^{-(X - N\langle x \rangle)^2 / 2N\sigma^2},\tag{2.16}$$

---

[4] Our presentation can be extended step-by-step to any spatial dimension.
[5] For brevity, we shall often write X instead of X_N.

where $\sigma^2 \equiv \langle x^2 \rangle - \langle x \rangle^2$. The conditions required for the central limit theorem to hold are: (i) the first two moments of the single-step distribution,

$$\langle x \rangle = \int_{-\infty}^{\infty} x \, p(x) \, dx \qquad \text{and} \qquad \langle x^2 \rangle = \int_{-\infty}^{\infty} x^2 \, p(x) \, dx \, ,$$

are finite,[6] (ii) the initial spatial distribution must also have finite first and second moments, and (iii) steps are independent. From the central limit theorem (2.16), the mean displacement and the variance are

$$\langle X_N \rangle = N \langle x \rangle \, , \qquad \text{var}(X_N) \equiv \langle X_N^2 \rangle - \langle X_N \rangle^2 = N \sigma^2 \, .$$

To derive the central limit theorem we notice that, since the steps of the random walk are independent, the probability distribution after N steps is related to the probability after $N - 1$ steps by the recursion (also known as the Chapman–Kolmogorov equation)

$$P_N(X) = \int_{-\infty}^{\infty} P_{N-1}(X') \, p(X - X') \, dX' \, . \tag{2.17}$$

Indeed, to reach X in N steps, the walk first reaches X' in $N-1$ steps and then makes a single step from X' to X with probability $p(X' \to X) = p(X - X')$. Equation (2.17) is a convolution and this suggests employing the Fourier transform to simplify the integral equation (2.17) to the recurrence $P_N(k) = P_{N-1}(k) \, p(k)$. Iterating yields $P_N(k) = P_0(k) \, [p(k)]^N$. For the canonical initial condition that the walk starts at the origin, $P_0(x) = \delta_{x,0}$, which gives $P_0(k) = 1$. Thus $P_N(k) = [p(k)]^N$ and

$$P_N(X) = \frac{1}{2\pi} \int_{-\infty}^{\infty} [p(k)]^N \, e^{-ikX} \, dk. \tag{2.18}$$

Because the first two moments of $p(x)$ are finite, we write the Fourier transform $p(k)$ as

$$p(k) = \int_{-\infty}^{\infty} p(x) \, e^{ikx} \, dx$$

$$= \int_{-\infty}^{\infty} p(x) \left[1 + ikx - \frac{1}{2} k^2 x^2 + \cdots \right] dx$$

$$= 1 + ik \langle x \rangle - \frac{1}{2} k^2 \langle x^2 \rangle + \cdots \, .$$

Then the asymptotic probability distribution is

$$P_N(X) \simeq \frac{1}{2\pi} \int_{-\infty}^{\infty} [1 + ik \langle x \rangle - \frac{1}{2} k^2 \langle x^2 \rangle]^N \, e^{-ikX} \, dk$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp \left[N \ln \left(1 + ik \langle x \rangle - \frac{1}{2} k^2 \langle x^2 \rangle \right) \right] e^{-ikX} \, dk$$

$$\simeq \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp \left[N \left(ik \langle x \rangle - \frac{k^2}{2} \left(\langle x^2 \rangle - \langle x \rangle^2 \right) \right) \right] e^{-ikX} \, dk. \tag{2.19}$$

---

[6] The central limit theorem continues to hold even if single-step moments beyond the second are divergent.

We complete the square in the exponent and perform the resulting Gaussian integral to arrive at the central limit theorem result (2.16).

2.3 Walks with broad distributions

What happens when the displacement distribution for a single step of a random walk is so wide that either its first moment and/or its second moment are divergent? In a similar vein, what happens if the time for each step is not fixed but drawn from a distribution whose first and/or second moments are divergent? When these single-event distributions are sufficiently broad, the distribution $P(X_N)$ is no longer Gaussian and the moments of the displacement scale anomalously with N. The entire walk may, in fact, be dominated by an exceptionally long single step or an exceptionally long hopping time that constitutes a finite fraction of the total. The existence of such exceptional events invalidates the conditions of the central limit theorem.

Broad step lengths

Suppose that the single-step distribution is given by

$$p(x) = \begin{cases} \mu\, x^{-(1+\mu)}, & x > 1, \\ 0, & 0 < x < 1, \end{cases} \tag{2.20}$$

with $\mu > 0$. We impose the lower cutoff to avoid complications due to infinitesimally short steps, while the condition $\mu > 0$ ensures a normalizable probability distribution. We now exploit basic facts about extreme value statistics (see the highlight below) to determine the first two moments of the displacement after N steps. The basic idea is to replace the true single-step distribution (2.20) by an *effective* distribution $p_{\text{eff}}(x)$ that describes a walk with a finite number of steps. This effective distribution has a bounded support whose upper limit is given by the largest step one can expect in a finite number of steps.

Extreme value statistics

Suppose that a sample of N random variables are independently drawn from the same continuous distribution $p(x)$ that is non-zero in the range $(0, \infty)$. The largest variable x_{max} in this sample is necessarily finite and we want to determine its statistical properties. An estimate for x_{max} is given by the *extremal criterion* that one of the N variables has a value that is greater than or equal to x_{max} (see Fig. 2.3):

$$\int_{x_{\text{max}}}^{\infty} p(x)\, dx \sim \frac{1}{N}. \tag{2.21}$$
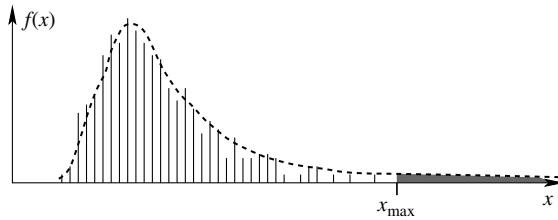
Schematic illustration of a finite sample of random variables (histogram) that are drawn from the continuous distribution $f(x)$. The tiny shaded area under the curve illustrates Eq. (2.21), which provides an estimate for $x_{\max}$ in this finite sample.

Continued

More properly, we determine the probability distribution that the largest out of N variables has a value x, from which the average value of $x_{\max}$ follows. For the largest variable to be equal to x, the remaining $N - 1$ variables must all be less than x, while one variable equals x. The probability for this event is given by

$$L_N(x) = N[1 - P(x)]^{N-1} p(x), \tag{2.22}$$

where $P(x) = \int_x^\infty p(y) \, dy$ is the probability that one variable lies in the range (x, ∞). In (2.22), the factor $[1 - P(x)]^{N-1}$ gives the probability that $N - 1$ variables are less than x and the factor of N appears because any of the N variables could be the largest. Since $P'(x) = -p(x)$, we have $\int_0^\infty L_N(x) \, dx = 1$, and L_N is properly normalized. For large N, the distribution $L_N(x)$ is approximately $Np(x)e^{-NP(x)}$, so the typical value of $x_{\max}$ can be found by requiring that the factor in the exponent is of the order of unity: $NP(x_{\max}) \sim 1$. This coincides with the rule-of-thumb estimate (2.21).

Using the extremal criterion of Eq. (2.21), we find $x_{\max} \sim N^{1/\mu}$. By construction, this length provides an upper cutoff on the single-step distribution for an N-step walk. Consequently, we should replace the single-step distribution $p(x)$ for an infinite sample by an *effective* single-step distribution for an N-step walk:

$$p_{\text{eff}}(x) = \begin{cases} \dfrac{\mu}{1 - x_{\max}^{-\mu}} \, x^{-(1+\mu)} \simeq \mu \, x^{-(1+\mu)}, & 1 < x < x_{\max}, \\[2mm] 0, & \text{otherwise.} \end{cases} \tag{2.23}$$

That is, the single-step distribution is cut off at $x_{\max}$, as larger steps hardly ever occur within N steps.

Example 2.1. *Probability distribution for the longest step*. This probability density $L_N(x)$ is given by Eq. (2.22) (see the highlight on the previous page). Since

$$P(x) = \int_x^\infty \mu \, y^{-(1+\mu)} \, dy = \begin{cases} x^{-\mu}, & x > 1, \\ 1, & 0 < x \le 1, \end{cases}$$

we have

$$L_N(x) = N(1 - x^{-\mu})^{N-1} \mu x^{-(1+\mu)}$$

for $x > 1$. Then the average length of the longest step is

$$\langle x_{\max} \rangle = \int_1^\infty x L_N(x)\, dx$$

$$= N \int_0^1 (1 - \xi)^{N-1} \xi^{-1/\mu}\, d\xi \qquad (\xi = x^{-\mu})$$

$$= \Gamma(1 - 1/\mu) \frac{\Gamma(N+1)}{\Gamma(N+1-1/\mu)}$$

$$\to \Gamma(1 - 1/\mu)\, N^{1/\mu} \quad \text{for} \quad N \gg 1.$$

Thus the naive extremal criterion (2.21) gives the correct N dependence of $\langle x_{\max} \rangle$; an exact approach is required for the computation of the amplitude.

The truncated single-step distribution $p_{\mathrm{eff}}(x)$ now satisfies the conditions of the central limit theorem – both $\langle x \rangle$ and $\langle x^2 \rangle$ are finite because of the cutoff. Thus we may compute the displacement of an N-step random walk with a broad distribution of individual step lengths by using $p_{\mathrm{eff}}(x)$ instead of the original single-step distribution $p(x)$ in the statements of the central limit theorem. Thus using $p_{\mathrm{eff}}(x)$, the mean length of a single step is given by

$$\langle x \rangle_{\mathrm{eff}} \sim \mu \int_1^{x_{\max}} x\, x^{-(1+\mu)}\, dx \sim \begin{cases} x_{\max}^{1-\mu}, & \mu < 1, \\ \ln x_{\max}, & \mu = 1, \\ \text{finite}, & \mu > 1. \end{cases} \tag{2.24a}$$

Since $x_{\max} \sim N^{1/\mu}$, the quantity $\langle x \rangle_{\mathrm{eff}}$ has the following N dependence:

$$\langle x \rangle_{\mathrm{eff}} \sim \begin{cases} N^{(1-\mu)/\mu}, & \mu < 1, \\ \ln N, & \mu = 1, \\ \text{finite}, & \mu > 1. \end{cases} \tag{2.24b}$$

Similarly $\langle x^2 \rangle_{\mathrm{eff}}$ is given by

$$\langle x^2 \rangle_{\mathrm{eff}} \sim \begin{cases} N^{(2-\mu)/\mu}, & \mu < 2, \\ \ln N, & \mu = 2, \\ \text{finite}, & \mu > 2. \end{cases} \tag{2.25}$$

Because the first two moments of the single-step distribution are finite, the mean displacement and the variance of an N-step random walk are, from the central limit theorem,

$$\langle X \rangle \simeq N \langle x \rangle_{\mathrm{eff}} \sim \begin{cases} N^{1/\mu}, & \mu < 1, \\ N \ln N, & \mu = 1, \\ N, & \mu > 1 \end{cases} \tag{2.26a}$$

and

$$\text{var}(X) = \langle X^2 \rangle - \langle X \rangle^2 \sim \begin{cases} N^{2/\mu}, & \mu < 2, \\ N \ln N, & \mu = 2, \\ N, & \mu > 2. \end{cases} \qquad (2.26b)$$

These results tell us that the most significant deviation from the Gaussian behavior occurs when $0 < \mu < 1$; this regime corresponds to so-called *Lévy flights*. The trajectory of a typical Lévy flight is shown in Fig. 2.1(b). This walk consists of many short segments and a few long segments that give the main contribution to the overall displacement; this behavior strongly contrasts with the trajectory of a random walk in Fig. 2.1(a), in which the length of each step is fixed.

When the tail of the single-step distribution decays more quickly than x^{-2} for large x, the probability distribution of the scaling variable $z = (X_N - N\langle x \rangle)/\sqrt{N}$ converges to a Gaussian form as $N \to \infty$. In the regime $0 < \mu < 2$, the probability distribution approaches the *Lévy distribution*[7] that is often written as L_μ. An important feature of this distribution is that it has a power-law tail, $L_\mu(z) \sim z^{-(1+\mu)}$. For $0 < \mu < 1$, the scaling variable $z = X_N/N^{1/\mu}$ obeys the Lévy distribution, while for $1 < \mu < 2$, the appropriate scaling variable is $z = (X_N - N\langle x \rangle)/N^{1/\mu}$.

Replacing the displacement X_N by the scaling variable z provides the proper way of taking the $N \to \infty$ limit. At the same time this replacement obscures a crucial difference between the Gaussian regime ($\mu > 2$) and Lévy flights ($0 < \mu < 1$). Indeed, in the Gaussian regime, the sum of a large number of independent identically distributed random variables becomes progressively more deterministic as $N \to \infty$. Namely, the sum is concentrated around $N\langle x \rangle$; although the deviations grow (on average) with N, the relative fluctuations vanish as $N^{-1/2}$ in the $N \to \infty$ limit. Thus randomness effectively disappears from the sum. This explains why we can trust large public opinion polls. This phenomenon is called *self-averaging*. For Lévy flights, however, the sum remains random. This randomness becomes evident if we notice that the dependence of $\langle X \rangle$ on N is the same as $\langle x_{\max} \rangle$! Thus Lévy flights are non-self-averaging, which is manifested by large sample-to-sample fluctuations that never disappear even in the thermodynamic limit.

In the above discussion we considered Lévy distributions that arise in the special case (2.20) when all steps are positive. More generally, the Lévy distribution $L_{\mu,\beta}$ depends on the index μ and on an asymmetry parameter β, with $-1 \leq \beta \leq 1$, that characterizes the relative frequency of positive steps ($\beta = 0$ corresponds to equal frequencies of positive steps and $\beta = 1$ corresponds to positive steps only). The Lévy distributions (2.20) that arise from the sums of random variables correspond to the fully asymmetric case $\beta = 1$; therefore $L_\mu \equiv L_{\mu,1}$. The symmetric Lévy distributions ($\beta = 0$) also often arise in applications. Their

---

[7] However, the ubiquity of the Gaussian distribution far exceeds that of all Lévy distributions. One very useful property that Lévy distributions share with the Gaussian is that these distributions are *stable*. By definition, a random variable is stable if the weighted sum of two copies of this variable has the same distribution as the original variable, apart from a possible shift and dilatation.

Fourier transforms are given by the elegant formula

$$L_{\mu,0}(z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{C|k|^{\mu}-ikz}\, dk. \tag{2.27}$$

The integral in (2.27) can be expressed via elementary functions in a few cases. The Gaussian distribution is recovered when $\mu = 2$, while for $\mu = 1$ the result is the Cauchy distribution[8]

$$L_{1,0}(z) = \frac{1}{\pi}\, \frac{1}{1+z^2}. \tag{2.28}$$

Various properties of walks with broad step lengths are the subject of problems 2.8–2.10.

Broad waiting times

Consider now the complementary situation where the time for each step is a random variable with a broad distribution. For the present discussion, the spatial nature of the single-step distribution is immaterial and we may use, for example, a nearest-neighbor random walk where the steps are equiprobably ± 1. When the average waiting time for a single step is finite, we anticipate normal behavior; that is, the mean-square displacement of a spatially unbiased walk should asymptotically grow as $\langle X^2 \rangle \sim t$. To see what can happen when the mean waiting time between steps is infinite, consider the waiting time distribution

$$\psi(\tau) = \begin{cases} \mu\,\tau^{-(1+\mu)}, & \tau > 1, \\ 0, & 0 < \tau < 1. \end{cases}$$

If $0 < \mu < 1$, the mean waiting time diverges, $\langle \tau \rangle = \infty$, and we anticipate anomalous behavior.

Proceeding exactly as in the case of broadly distributed step lengths, we make a correspondence to a random walk with a suitably truncated waiting time distribution. For $0 < \mu < 1$, the maximum waiting time over N steps is given by the extremal condition

$$\int_{\tau_{\max}}^{\infty} \mu\,\tau^{-(1+\mu)}\, d\tau \sim N^{-1},$$

from which the maximum waiting time for an N-step walk is $\tau_{\max} \sim N^{1/\mu}$. We now use the truncated waiting time distribution in which $\psi_{\mathrm{eff}}(\tau) = \psi(\tau)$ for τ in the range $[1, \tau_{\max}]$ and $\psi_{\mathrm{eff}}(\tau) = 0$ otherwise, to find that the average time for a random walk to take one step is

$$\langle t \rangle_{\mathrm{eff}} \sim \int_0^{\tau_{\max}} \mu\,\tau^{-\mu}\, d\tau \sim \begin{cases} N^{(1-\mu)/\mu}, & \mu < 1, \\ \ln N, & \mu = 1, \\ \text{finite}, & \mu > 1. \end{cases}$$

---

[8] The distribution in Eq. (2.28) is also known as the Lorentz distribution; in nuclear and particle physics it is often called the Breit–Wigner distribution.

Again, the subscript denotes that this average time pertains to the truncated distribution $\psi_{\text{eff}}(\tau)$. The total time T_N needed to take N steps is

$$T_N = N\langle t\rangle_{\text{eff}} \sim \begin{cases} N^{1/\mu}, & \mu < 1, \\ N\ln N, & \mu = 1, \\ N, & \mu > 1. \end{cases} \tag{2.29}$$

The mean-square displacement after N steps scales linearly with N and using (2.29) we infer the temporal scaling

$$\langle X^2\rangle \sim N \sim \begin{cases} T^\mu, & \mu < 1, \\ T/\ln T, & \mu = 1, \\ T, & \mu > 1. \end{cases} \tag{2.30}$$

The first two cases correspond to *subdiffusion*, where the mean-square displacement grows more slowly than linearly with time.

2.4 Application to gravity: the Holtsmark distribution

A physical realization of a random walk with a broad distribution of displacements arises in the distribution of gravitational fields – the *Holtsmark* distribution – that is generated by a random distribution of matter. Consider an infinite system of stars that are randomly distributed with a uniform density and with no correlations in spatial positions. We want to compute the distribution of the gravitational force acting on a single "test" star that we take to be at the origin with loss of generality. For simplicity, suppose that stars have equal masses (the general case of random star masses can also be treated, but the formulae become more cluttered). We are interested in the distribution of the random variable:

$$\mathbf{F} = \sum \mathbf{f}_j, \qquad \mathbf{f}_j = \frac{GM\mathbf{r}_j}{(r_j)^3}. \tag{2.31}$$

Here $\mathbf{r}_j$ is the location of the jth star, $\mathbf{f}_j$ is the force on the test star, G is Newton's gravitational constant, and M is the mass of each star. We ignore the factor GM henceforth; it can be easily restored in the final results on dimensional grounds.

We may interpret (2.31) as an infinite random walk with a broad distribution of (vector) steps $\mathbf{f}_j$. It is convenient to begin with a finite system and then take the thermodynamic limit. Specifically, consider a sphere of radius R with the center at the origin. The number of stars in this sphere is close to $N = nV$, where n is the density of stars and $V = \frac{4}{3}\pi R^3$. We first want to determine the *finite* sum $\mathbf{F} = \sum_{1\leq j\leq N} \mathbf{f}_j$.[9] Since the locations $\mathbf{r}_j$ of stars are random

---

[9] Strictly speaking, we should write the finite sum as $\mathbf{F}_N$, but we drop this subscript since it is understood by context.

and uniformly distributed within the sphere of volume V, the probability distribution $P(\mathbf{F})$ is given by the N-fold integral

$$P(\mathbf{F}) = \int \frac{d\mathbf{r}_1}{V} \times \cdots \times \frac{d\mathbf{r}_N}{V} \, \delta\left(\sum_{1 \leq j \leq N} \mathbf{f}_j - \mathbf{F}\right) . \tag{2.32}$$

Fourier-transforming, we obtain

$$P(\mathbf{k}) = \int d\mathbf{F} \, e^{i\mathbf{k}\cdot\mathbf{F}} \, P(\mathbf{F}) = \int \frac{d\mathbf{r}_1}{V} \times \cdots \times \frac{d\mathbf{r}_N}{V} \int d\mathbf{F} \, e^{i\mathbf{k}\cdot\mathbf{F}} \, \delta\left(\sum_{1 \leq j \leq N} \mathbf{f}_j - \mathbf{F}\right)$$

$$= \int \frac{d\mathbf{r}_1}{V} \times \cdots \times \frac{d\mathbf{r}_N}{V} \, e^{i\mathbf{k}\cdot\sum \mathbf{f}_j}$$

$$= \left[\int_{r<R} \frac{d\mathbf{r}}{V} \, e^{i\mathbf{k}\cdot\mathbf{f}}\right]^N , \qquad \mathbf{f} = \frac{\mathbf{r}}{r^3} .$$

We now rewrite the last integral as

$$\int_{r<R} \frac{d\mathbf{r}}{V} \, e^{i\mathbf{k}\cdot\mathbf{f}} = \int_{r<R} \frac{d\mathbf{r}}{V} \left[1 - \left(1 - e^{i\mathbf{k}\cdot\mathbf{f}}\right)\right] = 1 - \frac{1}{V} \int_{r<R} d\mathbf{r} \left(1 - e^{i\mathbf{k}\cdot\mathbf{f}}\right) , \tag{2.33}$$

to simplify the calculation of the Nth power. Indeed, we obtain

$$\left[\int_{r<R} \frac{d\mathbf{r}}{V} \, e^{i\mathbf{k}\cdot\mathbf{f}}\right]^N = \left[1 - \frac{1}{V} \int_{r<R} d\mathbf{r} \left(1 - e^{i\mathbf{k}\cdot\mathbf{f}}\right)\right]^N \to \exp\left[-n \int d\mathbf{r} \left(1 - e^{i\mathbf{k}\cdot\mathbf{f}}\right)\right] \tag{2.34}$$

in the thermodynamic limit $N, V \to \infty$, with $N/V = n$ constant. The last integral in (2.34) converges and therefore can be taken over all space rather than a sphere of radius R. From the inverse Fourier transform, the force distribution is given by

$$P(\mathbf{F}) = (2\pi)^{-3} \int d\mathbf{k} \, \exp[-i\mathbf{k} \cdot \mathbf{F} - n\Phi(\mathbf{k})] , \tag{2.35}$$

where we use the shorthand notation

$$\Phi(\mathbf{k}) = \int d\mathbf{r} \left(1 - e^{i\mathbf{k}\cdot\mathbf{f}}\right) , \qquad \mathbf{f} = \frac{\mathbf{r}}{r^3} . \tag{2.36}$$

To obtain explicit results we must now compute the integral (2.36) and then invert the Fourier transform (2.35). However, we can determine the dependence of $\Phi(\mathbf{k})$ on $\mathbf{k}$ without calculation. First, since the factor in the exponential $\mathbf{k} \cdot \mathbf{r}$ is integrated over all directions of $\mathbf{r}$, the result cannot depend on the direction of $\mathbf{k}$. Thus $\Phi(\mathbf{k}) = \Phi(k)$. Next, we can scale out the dependence of k in the exponential factor in (2.36) by defining a new integration variable $\mathbf{r} \to k^{-1/2}\mathbf{r}$. We thereby find that the overall k dependence of the integral is given by

$$\Phi(\mathbf{k}) = ak^{3/2}. \tag{2.37}$$

It is not possible to determine the amplitude a by symmetry considerations alone; however, a detailed calculation gives $a = \frac{4}{15} (2\pi)^{3/2}$ (problem 2.11).

Thus the force distribution is given by

$$P(\mathbf{F}) = (2\pi)^{-3} \int d\mathbf{k} \, \exp\left[-i\mathbf{k} \cdot \mathbf{F} - ank^{3/2}\right], \qquad a = \frac{4}{15} (2\pi)^{3/2}. \tag{2.38}$$

Using spherical coordinates in $\mathbf{k}$-space with axis along the direction of $\mathbf{F}$, we write $\mathbf{k} \cdot \mathbf{F} = kF \cos\theta$ and $d\mathbf{k} = 2\pi \sin\theta \, d\theta \, k^2 dk$. We then integrate Eq. (2.38) over the angular coordinate to give

$$\begin{aligned}
P(\mathbf{F}) &= (2\pi)^{-2} \int_0^\infty dk \, k^2 \int_0^\pi d\theta \, \sin\theta \, \exp\left[-ikF \cos\theta - ank^{3/2}\right] \\
&= (2\pi)^{-2} \int_0^\infty dk \, k^2 \, e^{-ank^{3/2}} \, 2 \, \frac{\sin kF}{kF} \\
&= \frac{1}{2\pi^2 F^3} \int_0^\infty dz \, z \, \sin z \, e^{-(z/\zeta)^{3/2}}, \quad \zeta = F/(an)^{2/3}.
\end{aligned} \tag{2.39}$$

To gain an understanding of the behavior of the force distribution, let us look at the limits of weak and strong forces. The integral in Eq. (2.39) depends on the combination $\zeta = F/(an)^{2/3}$, so the weak force limit corresponds to $\zeta \ll 1$, while $\zeta \gg 1$ gives the strong force limit.[10]

Example 2.2. *Weak force limit.* Since the dominant contribution to the integral in (2.39) comes from the region $z \sim \zeta$, the asymptotic behavior is easy to extract for $\zeta \ll 1$. We use $\sin z \simeq z$ to simplify (2.39) to

$$\begin{aligned}
P(\mathbf{F}) &\simeq \frac{1}{2\pi^2 F^3} \int_0^\infty dz \, z^2 \, e^{-(z/\zeta)^{3/2}} \\
&= \frac{\zeta^3}{2\pi^2 F^3} \frac{2}{3} \int_0^\infty dv \, v \, e^{-v} \qquad \left[z = \zeta v^{2/3}\right] \\
&= \frac{1}{3\pi^2} \frac{1}{(an)^2}.
\end{aligned} \tag{2.40}$$

Example 2.3. *Strong force limit.* When $\zeta \gg 1$, we write $\sin z$ as the imaginary part of e^{iz}, so that

$$P(\mathbf{F}) = \frac{1}{2\pi^2 F^3} \, \text{Im}[J(\zeta)], \qquad J(\zeta) = \int_0^\infty dz \, z \, e^{iz} \, e^{-(z/\zeta)^{3/2}}. \tag{2.41}$$

Using the substitution $z = iu$ we obtain

$$J(\zeta) = -\int du \, u \, e^{-u} \, \exp\left[\frac{1-i}{\sqrt{2}} \left(\frac{u}{\zeta}\right)^{3/2}\right],$$

---

[10] Note that the typical separation between stars is $\ell \sim n^{-1/3}$, so the typical force is $F_{\text{typ}} \sim \ell^{-2} \sim n^{2/3}$. Hence $\zeta \sim F/F_{\text{typ}}$; this explains why $\zeta \sim 1$ separates the weak and strong force regimes.

and therefore

$$\text{Im}[J(\zeta)] = \int du\, u \exp\left[-u + \frac{1}{\sqrt{2}}\left(\frac{u}{\zeta}\right)^{3/2}\right] \sin\left[\frac{1}{\sqrt{2}}\left(\frac{u}{\zeta}\right)^{3/2}\right].$$

The dominant contribution comes from the region $u \sim 1$. Here we can replace the sine function by its argument and also keep only the first term in the exponential to give

$$
\begin{aligned}
P(\mathbf{F}) &\simeq \frac{1}{2\pi^2 F^3} \int_0^\infty du\, u\, e^{-u}\, \frac{1}{\sqrt{2}}\left(\frac{u}{\zeta}\right)^{3/2} \\
&= \frac{1}{2\pi^2 F^3}\frac{\Gamma(7/2)}{\sqrt{2}\,\zeta^{3/2}} = \frac{15}{8}\frac{1}{(2\pi)^{3/2}}\frac{an}{F^{9/2}} = \frac{1}{2}\frac{n}{F^{9/2}},
\end{aligned}
\tag{2.42}
$$

where in the last line we use the value of the gamma function,

$$\Gamma\left(\frac{7}{2}\right) = \frac{5}{2}\cdot\frac{3}{2}\cdot\frac{1}{2}\cdot\Gamma\left(\frac{1}{2}\right) = \frac{15}{8}\sqrt{\pi},$$

and the amplitude a in Eq. (2.38).

Example 2.4. *Average force and moments of the force.* The asymptotic behaviors of the force distribution, Eqs (2.40) and (2.42), show that the νth moment of the force

$$\langle F^\nu\rangle \equiv \int F^\nu\, P(\mathbf{F})\, d\mathbf{F} = 4\pi \int_0^\infty F^{2+\nu}\, P(\mathbf{F})\, dF$$

exists only when $-3 < \nu < 3/2$. Using (2.39) and changing the integration variable from F to ζ we obtain

$$\langle F^\nu\rangle = C_\nu\,(an)^{2\nu/3}, \quad C_\nu = \frac{2}{\pi}\int_0^\infty dz\, z\, \sin z \int_0^\infty \frac{d\zeta}{\zeta}\,\zeta^\nu\, e^{-(z/\zeta)^{3/2}}.$$

The integrals in C_ν are computed using the properties of the gamma function:

$$
\begin{aligned}
C_\nu &= \frac{4}{3\pi}\int_0^\infty dz\, z\, \sin z \int_0^\infty \frac{dt}{t}\, t^{-2\nu/3}\, e^{-z^{3/2}t} \qquad (t = \zeta^{-3/2}) \\
&= \frac{4}{3\pi}\,\Gamma(-2\nu/3)\int_0^\infty dz\, z^{1+\nu}\, \sin z \\
&= (an)^{2\nu/3}\,\Gamma(1-2\nu/3)\,\Gamma(2+\nu)\,\frac{\sin(\pi\nu/2)}{\pi\nu/2}.
\end{aligned}
$$

The integral in the second line may be computed by writing $z = iy$, i.e. by treating z as a complex variable and moving the integration contour from the real axis $(0,\infty)$ to the imaginary axis $(0, i\infty)$.

In particular, the average force is

$$\langle F\rangle = C_1(an)^{2/3}, \qquad C_1 = \frac{4\,\Gamma(1/3)}{\pi} \doteq 3.41093.$$

It is instructive to derive the decay law (2.42) directly without using the Holtsmark distribution (2.38). Because strong gravitational forces are caused by nearby stars, let's first compute the probability distribution for the distance R between a given star and its closest neighbor. Since stars are distributed randomly with a uniform density n, the probability that the closest star is within the shell $(R, R + dR)$ is given by

$$\exp\left(-\frac{4\pi}{3} R^3 n\right) 4\pi R^2 n \, dR. \tag{2.43}$$

Here the exponential factor ensures that the sphere of radius R that is centered at a given star contains no other stars. The strong force limit corresponds to the case when the distance is small; more precisely $R^3 n \ll 1$. In this situation the probability distribution simplifies to $4\pi R^2 n$. Since the attraction to the closest star dominates, the force asymptotically equals $F = R^{-2}$. Using $4\pi R^2 n \, dR = \Pi(F) \, dF$ in conjunction with $F = R^{-2}$, we obtain $\Pi(F) = 2\pi n / F^{5/2}$ in the strong force limit. This result agrees with (2.42) if we recall that the force distribution is rotationally symmetric so that $\Pi(F) = 4\pi F^2 P(\mathbf{F})$.

2.5 First-passage properties

Our discussion thus far has focused on *where* a random walk wanders. Here we focus on the complementary question: *when* does a random walk *first* reach a given point? More generally, what is the *first-passage probability* $F(\mathbf{r}, t)$ for a random walk to *first* reach $\mathbf{r}$ at time t? First-passage phenomena underlie a wide range of phenomena where a stochastic variable first reaches a specified value. Examples include the evolution of chemical reactions, the firing of neurons in the brain, or the triggering of stock options. This section outlines several basic ideas of first-passage processes.

Transience and recurrence

Suppose that a random walk begins at the origin of an infinite lattice at $t = 0$. Does the random walker *eventually* return to its starting point? Remarkably, the answer depends only the spatial dimension d:

- For $d \leq 2$, a random walk surely returns to the origin – the walk is *recurrent*.
- For $d > 2$, there is a non-zero probability that the walk never returns – the walk is *transient*.

Moreover, if a random walk is recurrent it returns infinitely often, but the average return time is infinite!

To understand the criterion for transience and recurrence, consider a typical trajectory of a random walk. After time t the walk explores a spherical domain of radius $L \sim \sqrt{Dt}$, while the number of sites visited grows linearly in t. The density of visited sites is therefore $\rho \sim t / L^d \sim t^{1 - d/2}$. For $d > 2$, ρ decreases as t increases, so that some points within the domain of exploration may never be visited – transience. For $d < 2$, the density is an
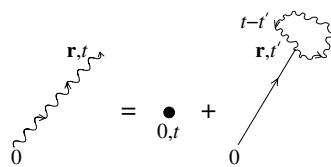
Fig. 2.4.   Diagrammatic representation of the relation between the occupation probability of a random walk (whose propagation is represented by a wavy line) and the first-passage probability (straight line).

increasing function of t, so that every site within the domain is surely visited – recurrence. The marginal case of $d = 2$ is barely recurrent, as we discuss below.

We now calculate the first-passage probability, from which the criterion for transience and recurrence can be inferred. The first-passage probability is related to the familiar occupation probability $P(\mathbf{r}, t)$ by

$$P(\mathbf{r}, t) = \int_0^t F(\mathbf{r}, t')\, P(\mathbf{0}, t - t')\, dt' + \delta_{\mathbf{r}, \mathbf{0}}\, \delta(t)\,, \tag{2.44}$$

where $\delta_{\mathbf{r}, \mathbf{0}}$ is the Kronecker delta function and $\delta(t)$ is the Dirac delta function. This equation accounts for the probability for a walk to reach $(\mathbf{r}, t)$ by a first passage to $\mathbf{r}$ at a time $t' \leq t$ followed by a loop $(\mathbf{r}, t') \to (\mathbf{r}, t)$ in the remaining time $t - t'$ (Fig. 2.4). Because the walk may also return to $\mathbf{r}$ at intermediate times between t' and t, the return factor involves the occupation probability P rather than F.

The integral on the right-hand side of Eq. (2.44) is a convolution. This suggests employing the Laplace transform.[11] In terms of the Laplace transforms of the occupation and first-passage probabilities, $P(\mathbf{r}, s) = \int_0^\infty P(\mathbf{r}, t)\, e^{-st}\, dt$ and $F(\mathbf{r}, s) = \int_0^\infty F(\mathbf{r}, t)\, e^{-st}\, dt$, Eq. (2.44) becomes the algebraic equation $P(\mathbf{r}, s) = F(\mathbf{r}, s)P(\mathbf{0}, s) + \delta_{\mathbf{r}, \mathbf{0}}$, from which

$$F(\mathbf{r}, s) = \frac{P(\mathbf{r}, s) - \delta_{\mathbf{r}, \mathbf{0}}}{P(\mathbf{0}, s)}\,. \tag{2.45}$$

The transience or recurrence of a random walk is determined by the *eventual* return probability to the origin:

$$\mathcal{R} = F(\mathbf{0}, 0) = \int_0^\infty F(\mathbf{0}, t)\, dt\,.$$

The walk is recurrent when $\mathcal{R} = 1$, while for $\mathcal{R} < 1$ the walk is transient. Since we only consider return to the origin, we drop the spatial argument in the following discussion.

---

[11] If a convolution is an integral with infinite limits, $\int_{-\infty}^\infty (\ldots)$, then we use the Fourier transform; this is what we've done in the case of the Chapman–Kolmogorov Eq. (2.17). When the limits in the convolution integral are finite, like in Eq. (2.44), the Laplace transform is preferable.

To compute the first-passage probability for the continuous-time random walk, we start with the generalization of Eq. (2.5) for the occupation probability at the origin on a d-dimensional hypercubic lattice:

$$P(t) = \left[I_0(2t)\, e^{-2t} \right]^d \simeq \frac{1}{(4\pi t)^{d/2}} \quad \text{when} \quad t \to \infty. \tag{2.46}$$

We use this basic result to compute the first-passage and eventual return probabilities for $d = 1$, 2, and 3.

Connection between algebraic tails of a function and its Laplace transform

Suppose that $f(t)$ has an algebraic tail, $f(t) \sim t^{-\mu}$ when $t \gg 1$, with $\mu < 1$, so that the integral $\int_0^\infty f(t)\, dt$ diverges. What is the Laplace transform $f(s)$? The answer obviously depends on the entire function $f(t)$, but the small-s asymptotic of the Laplace transform is governed by the large-t asymptotic of $f(t)$. By the definition of the Laplace transform,

$$f(s) = \int_0^\infty t^{-\mu}\, e^{-st}\, dt$$

for $f(t) = t^{-\mu}$. Making the substitution $x = st$ we rewrite the Laplace transform in the form

$$f(s) = s^{\mu-1} \int_0^\infty x^{-\mu}\, e^{-x}\, dx = \Gamma(1-\mu)\, s^{\mu-1} , \tag{2.47}$$

where we have used the definition of the gamma function,

$$\Gamma(z) = \int_0^\infty x^{z-1} e^{-x}\, dx .$$

Thus we obtain the handy relation between the tails of a function and its Laplace transform:

$$t^{-\mu} \quad \text{with} \quad \mu < 1 \quad \Longleftrightarrow \quad \Gamma(1-\mu)\, s^{\mu-1} , \tag{2.48}$$

which we will use extensively in this section and throughout this book.

The case $d = 1$. From Eqs (2.46) and (2.48), the Laplace transform of the occupation probability at the origin is $P(s) \simeq 1/\sqrt{4s}$. Then Eq. (2.45) gives, for the Laplace transform of the first-passage probability to the origin,

$$F(s) \simeq 1 - \sqrt{4s} . \tag{2.49}$$

Thus a random walk in one dimension is recurrent because the eventual return probability $\mathcal{R} = F(s = 0) = 1$. However, because $F(s)$ does not diverge for small s, it is not convenient to invert this Laplace transform directly. Instead, we consider the derivative of Eq. (2.49):

$$-\frac{dF(s)}{ds} = -\frac{d}{ds} \int_0^\infty F(t)\, e^{-st}\, dt = \int_0^\infty t\, F(t)\, e^{-st}\, dt = \frac{1}{\sqrt{s}} .$$

The point of this trick is that the Laplace transform of $tF(t)$ is more singular than the Laplace transform of $F(t)$ for small s, and we can then directly apply Eq. (2.48) to find that $t F(t) \simeq (\pi t)^{-1/2}$. Thus

$$F(t) \simeq \frac{1}{\sqrt{\pi}} \frac{1}{t^{3/2}} \quad \text{when} \quad t \to \infty. \tag{2.50}$$

The case $d = 2$. This case is subtle because the time integral of $P(t)$ diverges only weakly as $t \to \infty$. We must also cut off the integral for $t < 1$ to eliminate the spurious singularity that arises from using the long-time form of $P(t)$ for short times. With these caveats, we have

$$P(s) \simeq \int_1^\infty \frac{1}{4\pi t} e^{-st} \, dt = \frac{1}{4\pi} \int_s^\infty \frac{dx}{x} e^{-x} \simeq -\frac{1}{4\pi} \ln s. \tag{2.51}$$

The exponential factor in the integrand leads to a subdominant correction in the small-s limit, and we may ignore it to obtain the leading contribution to the integral. Using (2.51) in (2.45) we find

$$F(s) \simeq 1 + \frac{4\pi}{\ln s} \quad \text{when} \quad s \to 0. \tag{2.52}$$

A random walk is again recurrent in two dimensions because $F(s = 0) = \mathcal{R} = 1$. To solve for $F(t)$, we again apply $-d/ds$ to both sides of Eq. (2.52) to give

$$\int_0^\infty t F(t) e^{-st} \, dt \simeq \frac{4\pi}{s(\ln s)^2} \quad \text{when} \quad s \to 0. \tag{2.53}$$

Since the Laplace transform of a constant function equals $1/s$, the function $tF(t)$ must vary weakly in time to give the extra factor of $(\ln s)^{-2}$; namely (see problem 2.14)

$$F(t) \simeq \frac{4\pi}{t[\ln t]^2} \quad \text{when} \quad t \to \infty. \tag{2.54}$$

The case $d = 3$. This case is distinct from $d \le 2$ because

$$P(s) \simeq \int_1^\infty \frac{1}{(4\pi t)^{3/2}} e^{-st} \, dt \tag{2.55}$$

converges as $s \to 0$. Notice that the integral is again cut off for $t < 1$ to avoid the spurious singularity that arises from using the large-t form of $P(t)$ for all t. To determine $P(s)$ it is again useful to apply $-d/ds$ to both sides of Eq. (2.55) to give

$$-\frac{dP(s)}{ds} = \int_0^\infty \frac{1}{(4\pi)^{3/2}} \frac{1}{t^{1/2}} e^{-st} dt = \frac{1}{8\pi\sqrt{s}},$$

where we use Eq. (2.48) to evaluate the integral. Consequently $P(s)$ itself is given by

$$P(s) \simeq P(0) - \frac{\sqrt{s}}{4\pi}. \tag{2.56}$$

Then from Eq. (2.45), the Laplace transform of the first-passage probability is

$$F(s) \simeq [1 - P(0)^{-1}] - \frac{\sqrt{s}}{4\pi P(0)^2} = \mathcal{R} - \frac{(1-\mathcal{R})^2 \sqrt{s}}{4\pi} .$$

Because $P(s = 0) < \infty$, we conclude that $\mathcal{R} = [1 - P(0)^{-1}] < 1$ and the random walk is transient. Finally by comparing with Eq. (2.49), the asymptotic form of $F(t)$ is

$$F(t) \simeq \frac{(1 - \mathcal{R})^2}{8\pi^{3/2}} \frac{1}{t^{3/2}} . \tag{2.57}$$

2.6 Exit probabilities and exit times

Two of the most important first-passage properties are the *exit probability* and the *exit time*. Namely, for a diffusing particle that starts at a point $\mathbf{r}$ within a domain $\mathcal{B}$, what is the probability that the particle eventually reaches a point on the boundary and how long does it take for this event to occur? In this section, we present some basic features of these two quantities.

Exit probability

For the exit probability, it is natural to partition the boundary $\partial\mathcal{B}$ of a domain into two disjoint subsets $\partial\mathcal{B}_+$ and $\partial\mathcal{B}_-$, with $\partial\mathcal{B}_+ \cup \partial\mathcal{B}_- = \partial\mathcal{B}$, and ask for the probability $\mathcal{E}(\mathbf{r})$ that the particle exits the domain via $\partial\mathcal{B}_+$ before touching the boundary $\partial\mathcal{B}_-$. For example, in a finite one-dimensional interval $[0, N]$, $\partial\mathcal{B}_-$ would represent the point $x = 0$ and $\partial\mathcal{B}_+$ the point $x = N$, and we are interested, for example, in the probability that a walk that starts at an arbitrary point in the interior of the interval eventually reaches $x = N$ without ever touching $x = 0$.

Consider a discrete random walk that starts at site n on the interval $[0, N]$ and let $\mathcal{E}_n$ be the probability that the walk reaches N without visiting 0. This quantity satisfies the recurrence

$$\mathcal{E}_n = \frac{1}{2}(\mathcal{E}_{n-1} + \mathcal{E}_{n+1}) , \tag{2.58}$$

which asserts that, to exit starting from n, the walk either steps to the left with probability $\frac{1}{2}$, after which exit from $n - 1$ occurs, or steps to the right with probability $\frac{1}{2}$, after which exit from $n + 1$ occurs. The boundary conditions are $\mathcal{E}_0 = 0$ and $\mathcal{E}_N = 1$; i.e. if the walker starts at 0 it cannot exit at N, while it exits at N if it starts at this point. The recursion (2.58) is often referred to as a *backward* equation because it involves the starting point of the particle; in contrast, a master equation such as (2.4) for the probability distribution involves the current position of the particle.

The general solution to the discrete Laplace equation (2.58) is linear in the spatial variable, $A + Bn$, and invoking the boundary conditions gives

$$\mathcal{E}_n = \frac{n}{N}. \tag{2.59}$$

The reasoning that led to Eq. (2.58) applies to an arbitrary domain, and the exit probability in the continuum limit now satisfies the Laplace equation

$$\nabla^2 \mathcal{E}(\mathbf{r}) = 0, \qquad\qquad (2.60)$$

subject to the boundary conditions $\mathcal{E}(\mathbf{r} \in \partial \mathcal{B}_+) = 1$ and $\mathcal{E}(\mathbf{r} \in \partial \mathcal{B}_-) = 0$.

By this mapping of the exit probability into electrostatics, we can easily treat a variety of geometries. As a simple example, consider a diffusing particle that starts at a distance r from the center of a sphere of radius $R < r$. What is the probability that it will eventually hit the sphere? We must therefore solve $\nabla^2 \mathcal{E}(\mathbf{r}) = 0$ subject to the boundary condition $\mathcal{E}(r = R) = 1$. The answer depends fundamentally on the spatial dimension:

The case $d = 1$. The general solution to the Laplace equation is $\mathcal{E}(r) = A + Br$. Since $0 \le \mathcal{E}(r) \le 1$, B must be zero to ensure that the solution does not diverge as $r \to \infty$. Using the boundary condition $\mathcal{E}(r = R) = 1$ gives $\mathcal{E}(r) = 1$.

The case $d = 2$. The general solution to the Laplace equation is now $\mathcal{E}(r) = A + B \ln(r/R)$. The requirement $0 \le \mathcal{E}(r) \le 1$ again implies that $B = 0$, so the exit probability is once again trivial: $\mathcal{E}(r) = 1$.

The case $d > 2$. The general solution is $\mathcal{E}(r) = A + B r^{2-d}$, which no longer diverges as $r \to \infty$. The solution that obeys the boundary condition $\mathcal{E}(r) = 1$ is $\mathcal{E}(r) = 1 + B(r^{2-d} - R^{2-d})$. What has not yet been explicitly stated is that unbounded space is "unphysical"; we actually have in mind a problem with an external adsorbing boundary far from an internal sphere. Suppose that the external boundary is also a sphere of radius $R_\infty \gg R$ that is concentric with the internal sphere. Hence we solve the Laplace equation with the boundary conditions $\mathcal{E}(r = R) = 1$ and $\mathcal{E}(r = R_\infty) = 0$. The solution is

$$\mathcal{E}(r) = \frac{(R/r)^{d-2} - (R/R_\infty)^{d-2}}{1 - (R/R_\infty)^{d-2}}. \qquad\qquad (2.61)$$

Taking the $R/R_\infty \to 0$ limit we arrive at a Coulomb-like formula for the exit probability: $\mathcal{E}(r) = (R/r)^{d-2}$. It is plausible (although not easy to prove) that the final result is independent of the shape of the external surface as long as it is very far away from the internal sphere and we are seeking the exit probability $\mathcal{E}(\mathbf{r})$ for an initial point $\mathbf{r}$ that is far from the external surface.

To conclude, for $d \le 2$ a diffusing particle hits the internal sphere with certainty – this is a manifestation of recurrence. For $d > 2$, the probability that a diffusing particle that starts at r hits the internal sphere is just given by the Coulomb formula $\mathcal{E}(r) = (R/r)^{d-2}$.

Exit time

Let us begin by returning to a puzzling feature of the random walk in one dimension – the walker certainly visits the origin, yet the average time until the first visit is infinite. For concreteness, let's consider the related problem of continuum diffusion with diffusion

coefficient D. Suppose that the particle starts at $x = \ell > 0$. At a certain time T the particle will hit the origin for the first time; we call T the exit time. As we will show, the exit time is a random quantity with an infinite average, $\langle T \rangle = \infty$; nevertheless, the *typical* exit time is finite – it can depend only on D and ℓ – and dimensional analysis tells us that $T_{\text{typ}} \sim \ell^2/D$.

To understand this dichotomy, we compute the probability distribution $\Pi(T, \ell)$ of the exit time. A straightforward approach is to first calculate the probability density $P(x, t \mid \ell)$ that the walker starts at $x = \ell$, is at position x at time t, and has not reached the origin during the time interval $(0, t)$. Thus we solve the diffusion equation (2.7) subject to the initial condition $P(x, t = 0 \mid \ell) = \delta(x - \ell)$ and the boundary condition $P(x = 0, t \mid \ell) = 0$, which ensures that the particle has not reached the origin. The solution can be found using the image method, that is, by putting a fictitious antiparticle at $x = -\ell$ so that the initial condition becomes

$$P(x, t = 0 \mid \ell) = \delta(x - \ell) - \delta(x + \ell). \tag{2.62}$$

With this initial condition, the boundary condition $P(x = 0, t \mid \ell) = 0$ is manifestly obeyed. Owing to the linearity of the diffusion equation, the solution is just a combination of two Gaussian distributions,

$$P(x, t \mid \ell) = \frac{1}{\sqrt{4\pi Dt}} \left[e^{-(x-\ell)^2/4Dt} - e^{-(x+\ell)^2/4Dt} \right], \tag{2.63}$$

corresponding to a source at $x = \ell$ and sink at $x = -\ell$. The flux through the origin during the time interval $(t, t + dt)$ gives the probability density that the exit time is t (see problem 2.15). Thus

$$\Pi(T, \ell) = D \frac{\partial P}{\partial x} \bigg|_{x=0} = \frac{\ell}{\sqrt{4\pi DT^3}} e^{-\ell^2/4DT} . \tag{2.64}$$

As a formal definition of the typical time we use the quantity $\langle T^\alpha \rangle^{1/\alpha}$ where $\langle T^\alpha \rangle$ is the αth moment of the exit time

$$\langle T^\alpha \rangle = \int_0^\infty T^\alpha \, \Pi(T \mid \ell) \, dT.$$

Using (2.64) we obtain

$$\langle T^\alpha \rangle^{1/\alpha} = \frac{\ell^2}{4D} \left[\frac{\Gamma(\frac{1}{2} - \alpha)}{\Gamma(\frac{1}{2})} \right]^{1/\alpha} . \tag{2.65}$$

Typical times defined according to (2.65) are finite when $\alpha < 1/2$ and infinite otherwise; in particular, the average time $\langle T \rangle$ is infinite! However, when $\alpha < 1/2$, all moments $\langle T^\alpha \rangle^{1/\alpha}$ indeed scale as ℓ^2/D.

Consider now a diffusing particle in a finite domain. In such a domain, the average exit time is finite[12] and we can in principle compute the exit time distribution by employing the

---

[12] All other moments, e.g. $\langle T^2 \rangle, \langle T^3 \rangle$, etc., are also finite.

same scheme as in the case of the half-line. That is, we solve the diffusion equation

$$\frac{\partial P}{\partial t} = \nabla^2 P \tag{2.66}$$

inside domain $\mathcal{B}$ subject to the initial and the absorbing boundary conditions:

$$\begin{cases} P(\mathbf{x}, t = 0 \mid \mathbf{r}) = \delta(\mathbf{x} - \mathbf{r}), \\ P(\mathbf{x} \in \partial\mathcal{B}, t \mid \mathbf{r}) = 0. \end{cases} \tag{2.67}$$

The exit time distribution is given by the surface integral

$$\Pi(T, \mathbf{r}) = -D \int_{\partial\mathcal{B}} \nabla P(\mathbf{x}, T \mid \mathbf{r}) \cdot d\boldsymbol{\sigma} \tag{2.68}$$

and hence the average exit time (which is finite when the domain is bounded) is

$$t(\mathbf{r}) = \langle T(\mathbf{r}) \rangle = \int_0^\infty T \, \Pi(T, \mathbf{r}) \, dT. \tag{2.69}$$

Implementing the above procedure is laborious[13] whereas the final answers are often remarkably simple. If we are interested only in the average exit time, the above method for computing this time-*integrated* quantity is cumbersome because it requires (i) solving the time-dependent problem (2.66)–(2.67), (ii) integrating (2.68) to compute the exit time, and (iii) performing the average in (2.69).

We now describe a more elegant procedure that does not require analysis of time-dependent equations and the use of auxiliary quantities. We illustrate this approach for the continuous-time random walk on a one-dimensional lattice. By construction, in an infinitesimal time interval dt the exit time changes according to

$$T(x) = dt + \begin{cases} T(x), & \text{probability} \quad 1 - 2dt, \\ T(x + 1), & \text{probability} \quad dt, \\ T(x - 1), & \text{probability} \quad dt. \end{cases} \tag{2.70}$$

This equation is based on decomposing the random walk into the outcome after time dt and the remainder of the path. For example, if the walk hops from x to $x + 1$ in a time dt, then this walk will exit the system in a time dt (for the step just taken) plus $T(x + 1)$ (the exit time from the new location). The other two terms may be justified similarly. Averaging Eq. (2.70) yields

$$t(x) = dt + (1 - 2dt)t(x) + dt[t(x + 1) + t(x - 1)],$$

---

[13] For instance, when the domain is a finite interval $(0, L)$, this approach gives the probability density of exit times in the form of an infinite series. However, the average exit time for a particle that starts at $x \in (0, L)$ can be expressed in the simple form $t(x) = x(L - x)/2D$.

which simplifies to

$$t(x+1) - 2t(x) + t(x-1) = -1. \tag{2.71}$$

We now take the continuum limit of Eq. (2.71) by replacing the second difference by the second derivative, $t(x+1) - 2t(x) + t(x-1) \to D\,d^2t/dx^2$, with $D = 1$ for the continuous-time nearest-neighbor random walk in Fig. 2.2. Thus the average exit time satisfies

$$D\frac{d^2t}{dx^2} = -1 \tag{2.72}$$

in one dimension. Generally, the average exit time from an arbitrary domain obeys the Poisson equation

$$D\nabla^2 t(\mathbf{r}) = -1 \tag{2.73}$$

subject to the boundary condition $t(\mathbf{r} \in \partial\mathcal{B}) = 0$, which asserts that the exit time is zero if the diffusing particle starts on the domain boundary. As an example, consider the exit time to the surface of an absorbing sphere of radius R. Since all that matters is the radial distance, we only need the radial part of the Poisson equation

$$\frac{d^2t}{dr^2} + \frac{d-1}{r}\frac{dt}{dr} = -\frac{1}{D},$$

with solution

$$t(r) = \frac{R^2 - r^2}{2dD}. \tag{2.74}$$

The procedure for computing the exit time distribution is formidable, in principle, as it involves solving Eqs (2.66)–(2.67) for the density $P(\mathbf{x}, t \mid \mathbf{r})$, and then computing the spatial integral (2.68). The use of the Laplace transform allows us to avoid dealing with the probability density $P(\mathbf{x}, t \mid \mathbf{r})$, and instead work directly with the exit time distribution. Instead of giving a formal derivation, we derive the exit time distribution for the continuous-time nearest-neighbor random walk in one dimension and then take the continuum limit. The Laplace transform of the exit time distribution is

$$\Pi(s,x) = \int_0^\infty e^{-sT}\,\Pi(T,x)\,dT = \left\langle e^{-sT(x)} \right\rangle. \tag{2.75}$$

For the process defined by Eq. (2.70), the evolution equation for $\langle e^{-sT(x)} \rangle$ is

$$\left\langle e^{-sT(x)} \right\rangle = e^{-sdt}\left\{(1-2dt)\left\langle e^{-sT(x)}\right\rangle + dt\left\langle e^{-sT(x+1)}\right\rangle + dt\left\langle e^{-sT(x-1)}\right\rangle\right\}. \tag{2.76}$$

Expanding in powers of dt and taking the $dt \to 0$ limit we obtain

$$s\Pi(s,x) = \Pi(s,x+1) - 2\Pi(s,x) + \Pi(s,x-1),$$

which gives

$$s\Pi(s,x) = D\,\frac{d^2\Pi(s,x)}{dx^2} \tag{2.77}$$

in the continuum limit, and more generally

$$s\Pi(s,\mathbf{x}) = D\nabla^2\Pi(s,\mathbf{x}) \tag{2.78}$$
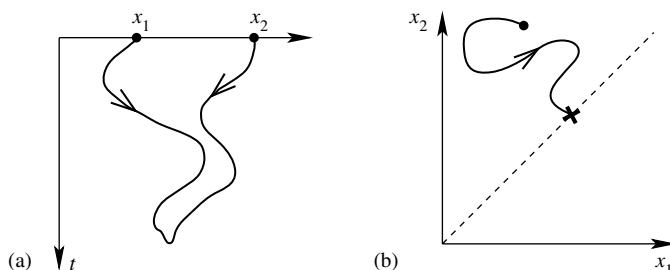
in higher dimensions. The exit time is equal to zero if the particle starts on the boundary, $\Pi(T,\mathbf{x}\in\partial\mathcal{B}) = \delta(T)$. Hence Eq. (2.78) should be supplemented by the boundary condition for the Laplace transform, $\Pi(s,\mathbf{x}\in\partial\mathcal{B}) = 1$.

Taken together, Eqs (2.60) and (2.78) provide a convenient and powerful approach to determine the exit probabilities and exit times for diffusing particles in arbitrary domains.

Application to vicious random walks

A natural application of first-passage processes is to the dynamics of mutually annihilating, or "vicious," random walks. What is their survival probability? We restrict ourselves to one dimension because the most interesting behavior occurs in this case. Let's start with the problem of two diffusing particles at x_1 and $x_2 > x_1$, with the same diffusion coefficient D, that both die when they meet. To find the survival probability, we first map this two-particle problem onto a single-particle problem by treating the coordinates x_1 and x_2 on the line as the position (x_1, x_2) of a single effective diffusing particle in two dimensions, subject to the constraint $x_2 > x_1$ (Fig. 2.5). If the effective particle hits the line $x_2 = x_1$ it dies. Because only the perpendicular distance to the line $x_2 = x_1$ matters, we have thus mapped the two-particle problem to that of a single diffusing particle at $y = x_2 - x_1$, with diffusion coefficient $2D$, and with an absorbing boundary at the origin.

This single-particle problem can be solved by the image method. We want the probability density for a diffusing particle on the line that starts at $y_0 > 0$, so that $c(y, t = 0) = \delta(y - y_0)$, and the particle dies if it reaches the origin. Here y_0 is the initial separation of the two particles

Fig. 2.5. Schematic illustration of the equivalence of two diffusing particles on the line (a) and a single diffusing particle in the region $x_2 > x_1$ (b). The former shows a space-time plot of the particle trajectories, while the latter shows (not to scale) the trajectory of the equivalent particle in the plane.

on the line. The probability density is (see also (2.63))

$$c(y, t) = \frac{1}{\sqrt{8\pi Dt}} \left[e^{-(y-y_0)^2/8Dt} - e^{-(y+y_0)^2/8Dt} \right] . \tag{2.79}$$

Since the initial condition is normalized, the first-passage probability to the origin at time t is just the flux to this point:

$$F(0, t) = 2D \left. \frac{\partial c(y, t)}{\partial y} \right|_{y=0} = \frac{y_0}{\sqrt{8\pi Dt^3}} \, e^{-y_0^2/8Dt} \sim t^{-3/2} \tag{2.80}$$

as $t \to \infty$, in agreement with the time dependence in Eq. (2.50). The survival probability $S(t)$ of the two particles may be found from $S(t) = 1 - \int_0^t F(0, t') \, dt'$. Introducing the variable $u^2 = y_0^2/8Dt'$ leads to

$$S(t) = 1 - \frac{2}{\sqrt{\pi}} \int_{y_0/\sqrt{8Dt}}^{\infty} e^{-u^2} \, du = \mathrm{erf}\left(\frac{y_0}{\sqrt{8Dt}} \right) \to \frac{y_0}{\sqrt{2\pi Dt}} \tag{2.81}$$

as $t \to \infty$. As expected by the recurrence of diffusion, the survival probability ultimately decays to zero. The two particles certainly annihilate; however, the average annihilation time, $\langle t \rangle \equiv \int_0^\infty t F(0, t) \, dt$, is infinite.

What happens with three particles at x_1, $x_2 > x_1$, and $x_3 > x_2$ on the line? If they are mutually vicious, the probability that they all survive until time t equals the probability that they all maintain the ordering $x_3 > x_2 > x_1$ for all times $t' \leq t$. It is again convenient to solve this problem by mapping to a single diffusing particle in three dimensions whose coordinates always satisfy $x_3 > x_2 > x_1$. The constraint $x_2 > x_1$ corresponds to the diffusing particle always remaining on one side of the plane $x_2 = x_1$. Similarly, the constraint $x_3 > x_2$ corresponds to the diffusing particle always remaining on one side of the plane $x_3 = x_2$. To satisfy both constraints, the particle must remain within the infinite two-dimensional wedge of opening angle $\pi/3$ that is defined by these two constraints. The survival probability of a diffusing particle inside an infinite two-dimensional wedge of opening angle ϕ asymptotically decays as

$$S(t) \sim t^{-\pi/2\phi} . \tag{2.82}$$

Consequently, the survival probability of three mutually vicious walkers asymptotically decays as $t^{-3/2}$.

The three-particle system contains more subtlety. Let's designate one of the particles as a "prey" and the other two as "predators" that do not interact with each other, but each kill the prey if they meet it. Assume that both species diffuse with the same diffusion coefficient. What is the prey survival probability? There are two distinct cases: (i) prey initially in the middle, and (ii) prey at the end. These three-particle problems can each be solved by mapping to a single diffusing particle in three dimensions subject to appropriate constraints. For the prey in the middle, the constraints $x_2 > x_1$ and $x_3 > x_2$ are the same as that for three mutually vicious walks and the prey survival probability asymptotically decays as $t^{-3/2}$.

For the prey at x_1 at one end, the survival constraints are now: $x_1 < x_2$ and $x_1 < x_3$. By mapping to the equivalent three-dimensional problem, the opening angle of the wedge is now $2\pi/3$, so that the prey survival probability asymptotically decays as $t^{-3/4}$.
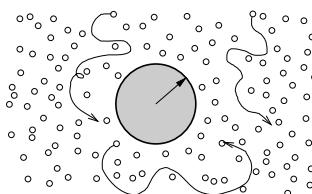
Notice the diminishing return of adding a second predator at the end: for one predator the survival probability decays as $t^{-1/2}$, while for two predators the survival probability decays slower than $(t^{-1/2})^2$. The natural next question is: what is the prey survival probability $S_N(t)$ for $N > 2$ predators on one side of the prey? The answer is unknown! The asymptotic decay is clearly algebraic, $S_N \sim t^{-\gamma_N}$. Numerically, it appears that $\gamma_3 \approx 0.91342$, $\gamma_4 \approx 1.032$, and the exponents γ_N slowly grow with N, e.g. $\gamma_{10} \approx 1.4$; for $N \to \infty$, a qualitative argument gives $\gamma_N \simeq \frac{1}{4} \ln 4N$.

2.7 Reaction rate theory

What is the rate at which diffusing molecules hit an absorbing object? The answer to this question sheds light on diffusion-controlled reactions, a subject that will be treated in Chapter 13. In such processes, a reaction occurs as soon as two reactants approach to within an interaction radius – effectively, the reactants "meet." The evolution of the reaction is therefore limited by the rate at which diffusion brings reactants in proximity. The goal of reaction rate theory is to calculate this rate.

To provide context for reaction rate theory, consider the much simpler example of the reaction rate when external particles move ballistically rather than diffusively. For a uniform beam of particles that is incident on an absorbing object, the reaction rate is clearly proportional to its cross-sectional area. In stark contrast, when the "beam" consists of diffusing particles, the reaction rate grows *more slowly* than the cross-sectional area and also has a non-trivial shape dependence. By employing the mathematical similarity between diffusion and electrostatics we will show how to determine the reaction rate of an object in terms of its electrical capacitance.

We first study the reaction rate of an arbitrarily shaped absorbing object in three dimensions ($d = 3$). The object is surrounded by a gas of non-interacting molecules, each of which is absorbed whenever it hits the surface of the object (Fig. 2.6). The reaction rate is defined as the steady-state diffusive flux to the object. To find this flux, we need to solve the diffusion equation for the concentration $c(\mathbf{r}, t)$ exterior to the object $\mathcal{B}$, with absorption on its boundary $\partial\mathcal{B}$. It is conventional to choose a spatially uniform unit density for the

Fig. 2.6. Schematic concentration profile of independent diffusing particles around an absorbing sphere.

initial condition. Thus we need to solve:

$$\frac{\partial c}{\partial t} = D\nabla^2 c, \qquad \text{with} \quad c(\mathbf{r} \in \partial\mathcal{B}, t) = 0, \quad c(\mathbf{r}, t = 0) = 1. \qquad (2.83)$$

An important simplifying feature for $d = 3$ is that the concentration approaches a steady state. Intuitively, this steady state arises because of the transience of diffusion, so that the loss of molecules by absorption is balanced by replenishment from afar. On the other hand, diffusion is recurrent for $d = 2$ and $d = 1$. Because a diffusing particle must eventually hit the absorber, there can be no steady-state solution to Eq. (2.83).

Because $c(\mathbf{r}, t)$ approaches the steady state $c_\infty(\mathbf{r})$ for $d = 3$, we study the complementary function $\phi(\mathbf{r}) = 1 - c_\infty(\mathbf{r})$. The governing equations for $\phi(\mathbf{r})$ are

$$D\nabla^2 \phi = 0, \quad \phi(\mathbf{r} \in \partial\mathcal{B}) = 1, \quad \phi(\mathbf{r} \to \infty) = 0. \qquad (2.84)$$

Since $\phi = 1$ on the boundary and $\phi(\mathbf{r})$ satisfies the Laplace equation, ϕ is just the *electrostatic potential* generated by a perfectly conducting object $\mathcal{B}$ that is held at unit potential.

By definition, the reaction rate K is given by

$$K = D\int_{\partial\mathcal{B}} \nabla c \cdot d\boldsymbol{\sigma} = -D\int_{\partial\mathcal{B}} \nabla\phi \cdot d\boldsymbol{\sigma} . \qquad (2.85)$$

On the other hand, according to electrostatics the total charge on the surface of the equivalent conductor in three dimensions is

$$Q = -\frac{1}{4\pi}\int_{\partial\mathcal{B}} \nabla\phi \cdot d\boldsymbol{\sigma} . \qquad (2.86)$$

Moreover, the total charge on the conductor is related to its capacitance C by $Q = C\phi_\mathcal{B}$. Consequently, when the conductor is held at unit potential the reaction rate is given by

$$K = 4\pi DQ = 4\pi DC . \qquad (2.87)$$

This fundamental equivalence allows us to find the reaction rate of various simple objects from known values of their capacitances.

- Sphere of radius R in three dimensions:[14]

$$C = R, \qquad K = 4\pi DR.$$

- Prolate ellipsoid of revolution with axes $a > b = c$:

$$C = \frac{\sqrt{a^2 - b^2}}{\cosh^{-1}(a/b)}, \qquad K = 4\pi D\frac{\sqrt{a^2 - b^2}}{\cosh^{-1}(a/b)}.$$

---

[14] For the sphere in $d > 2$ dimensions, $C = (d - 2)R^{d-2}$ and consequently $K = D(d - 2)\Omega_d R^{d-2}$ where $\Omega_d = 2\pi^{d/2}/\Gamma(d/2)$ is the surface area of a d-dimensional unit sphere.

- Oblate ellipsoid of revolution with axes $a = b > c$:

$$C = \frac{\sqrt{a^2 - c^2}}{\cos^{-1}(c/a)}, \qquad K = 4\pi D \frac{\sqrt{a^2 - c^2}}{\cos^{-1}(c/a)}.$$

- Disk of radius R:

$$C = 2R/\pi, \qquad K = 8DR.$$

What about the reaction rate in low-dimensional systems with $d \leq 2$? Because diffusion is recurrent, a continuously growing depletion zone develops around the absorber. As a result, the absorbed flux monotonically decreases to zero as $t \to \infty$. While the reaction rate defined by Eq. (2.85) approaches zero as $t \to \infty$, it is useful to think of a *time-dependent* reaction rate. To obtain this reaction rate, we now should solve the full time-dependent diffusion equation and then compute the flux to the absorber. While this calculation is tractable (problem 2.23), we can avoid most technical details by performing an approximate *quasi-static* calculation that we illustrate for an absorbing sphere. This approach is a powerful, yet simple method for solving the diffusion equation in the presence of slowly moving boundaries.

The basis of this quasi-static approximation is that the region exterior to the absorbing sphere naturally divides, for $d \leq 2$, into a "near" zone that extends a distance $\sqrt{Dt}$ from the surface, and the complementary "far" zone (Fig. 2.7). The near zone is the region in which diffusing particles are affected by the absorbing boundary. Nevertheless, diffusing particles still have ample time to explore this near zone thoroughly before being absorbed. Consequently, the concentration is nearly time independent. Conversely, in the far zone the probability of a particle being adsorbed is negligible because particles are unlikely to diffuse a distance greater that $\sqrt{Dt}$ during a time t. Thus the far-zone concentration remains close to its initial value, $c(r) \simeq 1$ for $r > \sqrt{Dt}$.

Based on this intuition, we solve the Laplace equation in the near zone, but with the *time-dependent* boundary condition $c(r = \sqrt{Dt}) = 1$ to match to the static far-zone solution at $r = \sqrt{Dt}$. The general solution to the Laplace equation has the form $c(r) = A + Br^{2-d}$ for $d < 2$ and $c(r) = A + B \ln r$ for $d = 2$. Matching to the boundary conditions at $r = R$ and
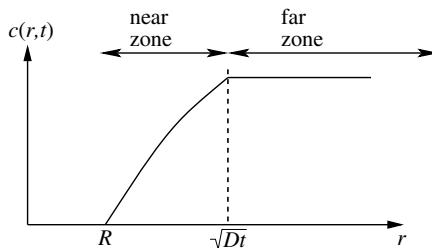
Fig. 2.7.　Sketch of the concentration about an absorbing sphere in the quasi-static approximation. The near- and far-zone concentrations match at $r = R + \sqrt{Dt}$.

at $r = \sqrt{Dt}$ for $t \to \infty$ gives

$$c(r,t) \simeq \begin{cases} \dfrac{1 - (R/r)^{d-2}}{1 - (R/\sqrt{Dt})^{d-2}} \to \left(\dfrac{\sqrt{Dt}}{r}\right)^{d-2}, & d < 2, \\[2ex] \dfrac{\ln(r/R)}{\ln(\sqrt{Dt}/R)}, & d = 2. \end{cases} \qquad (2.88)$$

Finally, we compute the flux $D \int_{\partial \mathcal{B}} \nabla c \cdot d\boldsymbol{\sigma}$ to the surface, from which we obtain the dimension dependence of the reaction rate to a sphere of radius R:

$$K(t) \sim \begin{cases} D \times (Dt)^{(d-2)/2}, & d < 2, \\[2ex] \dfrac{4\pi D}{\ln\left(Dt/R^2\right)}, & d = 2, \\[2ex] DR^{d-2}, & d > 2. \end{cases} \qquad (2.89)$$

The rate is time independent and grows with radius as R^{d-2} for $d > 2$; for $d \leq 2$ the rate is *independent* of the sphere radius and decreases as a power law in time when $d < 2$.

2.8 The Langevin approach

As mentioned at the outset of this chapter, the apparently random motion of Brownian particles is caused by a huge number of collisions with fluid molecules. There is no hope, and no point, in accounting for all of these collisions and thereby integrating the equation of motion exactly. The Langevin approach is based on replacing the effect of these very frequent collisions by an effective stochastic external force. In this way a many-body problem is reduced to a much more tractable effective single-body problem.

To illustrate the Langevin approach, let's begin with diffusive motion in one dimension.[15] We mimic the effect of collisions by a stochastic noise $\xi(t)$. The equation of motion for the position $x(t)$ of the particle,

$$\frac{dx}{dt} = \xi(t), \qquad (2.90)$$

is a *stochastic differential equation* because $\xi(t)$ is a stochastic variable that is different for each realization of the noise. This noise has the following basic properties:

- $\xi(t)$ is independent of x.
- $\xi(t)$ fluctuates on microscopically short time scales to mimic the effect of collisions.[16]

---

[15] The same treatment works in higher dimensions; we focus on one dimension for simplicity.

[16] For pollen grains, the objects of the experiments by Robert Brown, their linear dimension is of the order of 10^{-5} m and there are of the order of 10^{20} collisions per second. It is clearly impossible (and pointless!) to follow the collisional changes on this incredibly short 10^{-20} s time scale.

- $\langle \xi(t) \rangle = 0$ so that collisions do not conspire to give a net velocity (the bracket $\langle \cdots \rangle$ denotes averaging over different realizations of noise).
- There is no correlation between the noise at different times, $\langle \xi(t)\xi(t') \rangle = 2D\delta(t - t')$; we will justify the amplitude $2D$ subsequently.

Integrating Eq. (2.90) yields the formal solution

$$x(t) = \int_0^t \xi(t')\, dt'. \qquad (2.91)$$

When we average (2.91) over the noise we obtain $\langle x(t) \rangle = 0$ because $\langle \xi(t) \rangle = 0$. Squaring Eq. (2.91) gives

$$\langle x^2 \rangle = \int_0^t \int_0^t \langle \xi(t')\xi(t'') \rangle\, dt'\, dt''. \qquad (2.92)$$

Using $\langle \xi(t')\xi(t'') \rangle = 2D\delta(t' - t'')$, we find $\langle x^2 \rangle = 2Dt$, which reproduces the result from the diffusion equation. This connection with diffusion justifies the amplitude $2D$ in the noise correlation function. Notice that the time dependence follows from Eq. (2.90) by dimensional analysis. Since $\delta(t)$ has units of $1/t$ (because the integral $\int \delta(t)\, dt = 1$), the statement $\langle \xi(t)\xi(t') \rangle = 2D\delta(t - t')$ means that ξ has the units $\sqrt{D/t}$. Thus from Eq. (2.91) we see that $x(t)$ must have units of $\sqrt{Dt}$.

Not only is the variance identical to the prediction of the diffusion equation, the entire probability distribution is Gaussian. This fact can be seen by dividing the interval $(0, t)$ into the large number $t/\Delta t$ of sub-intervals of duration Δt, replacing the integral in Eq. (2.91) by the sum

$$x(t) = \sum_{j=1}^{t/\Delta t} \zeta_j, \quad \zeta_j = \int_{(j-1)\Delta t}^{j\Delta t} \xi(t')\, dt', \qquad (2.93)$$

and noting that the displacements ζ_j are independent identically distributed random variables satisfying $\langle \zeta_j \rangle = 0$, $\langle \zeta_j^2 \rangle = 2D\Delta t$. Therefore we apply the central limit theorem to obtain

$$P(x, t) = \frac{1}{\sqrt{4\pi Dt}}\, e^{-x^2/4Dt}. \qquad (2.94)$$

The Gaussian probability distribution is particularly striking in view of the feature that the Langevin equation (2.90) builds in an unphysical infinite velocity. Indeed, in a small time interval Δt, the mean-square displacement of a Brownian particle is $\langle \Delta x^2 \rangle = 2D\Delta t$. Therefore the typical velocity $v_{\text{typ}} \sim \Delta x/\Delta t \sim \sqrt{D/\Delta t}$ diverges as $\Delta t \to 0$. Mathematically, the function $x(t)$ is continuous but nowhere differentiable. Langevin proposed the equation

$$\frac{d\mathbf{v}}{dt} = -\gamma \mathbf{v} + \boldsymbol{\xi}(t), \qquad \frac{d\mathbf{x}}{dt} = \mathbf{v}, \qquad (2.95)$$

with the velocity $\mathbf{v}$, which remains finite, as the basic variable.[17] The Langevin approach (2.95) posits that the influence of the fluid can be divided into a systematic frictional force, as represented by the term $-\gamma\mathbf{v}$, and a fluctuating term represented by $\boldsymbol{\xi}(t)$. The frictional force is normally governed by Stokes' law[18] in which $\gamma = 6\pi a\eta/m$, where a is the particle radius (assumed spherical), η the viscosity coefficient of the fluid, and m the mass of the Brownian particle. The noise has zero mean value and the correlation function $\langle \xi_i(t)\xi_j(t')\rangle = \Gamma\delta_{ij}\delta(t - t')$. Since both these contributions are caused by the surrounding fluid, they are not independent; we shall see that γ and Γ are connected by a *fluctuation–dissipation relation*.

The formal solution to Eq. (2.95) is

$$\mathbf{v}(t) = \mathbf{v}_0\, e^{-\gamma t} + e^{-\gamma t}\int_0^t \boldsymbol{\xi}(t')\, e^{\gamma t'}\, dt'\,. \tag{2.96}$$

Therefore the average velocity decays exponentially: $\langle\mathbf{v}\rangle = \mathbf{v}_0\, e^{-\gamma t}$. The mean-square velocity exhibits a more interesting behavior:

$$\langle\mathbf{v}^2\rangle = \mathbf{v}_0^2\, e^{-2\gamma t} + e^{-2\gamma t}\int_0^t\int_0^t \langle\boldsymbol{\xi}(t')\,\boldsymbol{\xi}(t'')\rangle e^{\gamma(t'+t'')}\, dt'\, dt''$$

$$= \mathbf{v}_0^2\, e^{-2\gamma t} + \frac{3\Gamma}{2\gamma}\left(1 - e^{-2\gamma t}\right)\,. \tag{2.97}$$

The equipartition theorem tells us that in equilibrium each degree of freedom has an average energy $T/2$. (We always measure the temperature in energy units; equivalently, we set Boltzmann's constant $k = 1$.) Hence $\langle\mathbf{v}^2\rangle = 3T/m$ in three dimensions. By comparing with Eq. (2.97) in the limit $t \to \infty$, the friction coefficient and the noise amplitude are therefore related by the fluctuation–dissipation relation $\Gamma = T\gamma/m$.

To determine the probability distribution $P(\mathbf{v}, t\,|\,\mathbf{v}_0)$ from (2.96) we can proceed as in the derivation of (2.94) from (2.93), that is, we write

$$\mathbf{v}(t) - \mathbf{v}_0\, e^{-\gamma t} = \sum_{j=1}^{t/\Delta t} \boldsymbol{\zeta}_j\,, \quad \boldsymbol{\zeta}_j = e^{-\gamma t}\int_{(j-1)\Delta t}^{j\Delta t} \boldsymbol{\xi}(t')\, e^{\gamma t'}\, dt'\,. \tag{2.98}$$

On the right-hand side we have the sum of independent (albeit not identically distributed) random variables. The central limit theorem still holds in this case,[19] however, and taking the $\Delta t \to 0$ limit, we arrive at

$$P(\mathbf{v}, t\,|\,\mathbf{v}_0) = \left[\frac{m}{2\pi T(1 - e^{-2\gamma t})}\right]^{3/2} \exp\left[-\frac{m(\mathbf{v} - \mathbf{v}_0\, e^{-\gamma t})^2}{2T(1 - e^{-2\gamma t})}\right]\,, \tag{2.99}$$

which approaches the Maxwellian distribution in the $t \to \infty$ limit.

---

[17] From Eq. (2.95), the velocity is continuous but nowhere differentiable, so the acceleration is infinite. Note also that in the $\gamma \to \infty$ limit, we can ignore the inertia term $\dot{\mathbf{v}}$ and recover (2.90).

[18] The drag force is linear in velocity when velocity is small, or fluid is very viscous. More precisely, the dimensionless Reynolds number (that gives the ratio of inertial to viscous terms) should be small. For a sphere, the Reynolds number is $Re = \rho v a/\eta$.

[19] The full variance is the sum of individuals variances $\sum_{j=1}^{t/\Delta t}\langle\zeta_j^2\rangle$.

Our original goal was to describe the probability distribution of the displacement, so let's now examine this quantity. Integrating $\dot{\mathbf{x}} = \mathbf{v}$ we obtain

$$\mathbf{x}(t) = \mathbf{x}_0 + \gamma^{-1}\mathbf{v}_0(1 - e^{-\gamma t}) + \gamma^{-1}\int_0^t \boldsymbol{\xi}(t')\left[1 - e^{-\gamma(t-t')}\right]dt', \qquad (2.100)$$

so that the average displacement is $\langle \mathbf{x} \rangle = \mathbf{x}_0 + \gamma^{-1}\mathbf{v}_0(1 - e^{-\gamma t})$. Squaring (2.100) we find the mean-square displacement

$$\langle \mathbf{x}^2 \rangle = \mathbf{x}_0^2 + \gamma^{-2}\mathbf{v}_0^2\, e^{-2\gamma t} + 2\gamma^{-1}(1 - e^{-\gamma t})\mathbf{x}_0 \cdot \mathbf{v}_0 + \frac{3\Gamma}{2\gamma^3}\, g(t),$$

$$g(t) = 2\gamma t - 3 + 4e^{-\gamma t} - e^{-2\gamma t}. \qquad (2.101)$$

With a bit more work we obtain the entire distribution

$$P(\mathbf{x}, \mathbf{v}, t \mid \mathbf{x}_0, \mathbf{v}_0) = \left[\frac{m\gamma^2}{2\pi T g(t)}\right]^{3/2} \exp\left[-\frac{m\gamma^2[\mathbf{x} - \mathbf{x}_0 - \mathbf{v}_0(1 - e^{-\gamma t})/\gamma]^2}{2T g(t)}\right]. \qquad (2.102)$$

Asymptotically this probability distribution agrees with the solution to the diffusion equation when the diffusion constant is given by

$$D = \frac{T}{m\gamma} = \frac{T}{6\pi\eta a}. \qquad (2.103)$$

This connection between parameters is known as the *Stokes–Einstein* relation.

There exists a standard prescription to go from a Langevin equation for a stochastic variable x to the *Fokker–Planck* equation for the probability distribution of this variable $P(x, t)$. Because this connection is extensively discussed in standard texts, we do not treat this subject here. We merely point out that for the generic Langevin equation

$$\frac{dx}{dt} = F(x) + \xi(t),$$

in which the noise $\xi(t)$ has zero mean and no correlations, $\langle \xi \rangle = 0$ and $\langle \xi(t)\xi(t') \rangle = 2\Gamma\delta(t - t')$, the corresponding Fokker–Planck equation for $P(x, t)$ is

$$\frac{\partial P}{\partial t} = -\frac{\partial}{\partial x}[F(x)\, P] + \Gamma\frac{\partial^2 P}{\partial x^2}. \qquad (2.104)$$

For example, for the Langevin equation $\dot{x} = \xi(t)$, the corresponding Fokker–Planck equation reduces to the standard diffusion equation (2.7). The salient feature of the Fokker–Planck equation is that its solution always has a Gaussian form. We will exploit this property in our discussion in Section 12.3 about the small-fluctuation expansion of the master equation for reacting systems.

2.9 Application to surface growth

We have seen that a combination of the simplest noise and the most trivial deterministic differential equation $\dot{x} = 0$ leads to the Langevin equation (2.90), a remarkable

stochastic differential equation that contains all the properties of one-dimensional Brownian motion. Langevin equations corresponding to more complicated differential equations (e.g. to the harmonic oscillator) and more complicated noise terms have been subsequently investigated. The idea is always the same; we want to understand how the behavior of a deterministic system changes when it interacts with its environment. We cannot describe the environment in a precise manner and instead we mimic its influence on the system as a stochastic force.

Generally Langevin equations are constructed by the following standard procedure:

1. Start with a deterministic equation. Usually it is an ordinary differential equation, but other choices (e.g. a difference equation) are also possible.
2. Add a noise term. In the simplest scenario, the noise is assumed to be independent of the underlying variable, temporarily uncorrelated, and Gaussian.

Numerous natural phenomena are governed by *partial* differential equations. An ambitious generalization of the Langevin program is to start with a deterministic partial differential equation. The most important *linear* partial differential equation that appears in this book is the diffusion equation. The most prominent nonlinear generalization to the diffusion equation is the Burgers equation. Let us now look at Langevin equations corresponding to these two partial differential equations.

Edwards–Wilkinson equation

The Edwards–Wilkinson (EW) equation is the diffusion equation with noise:

$$\frac{\partial h}{\partial t} = D\nabla^2 h + \eta. \tag{2.105}$$

This Langevin equation has been proposed as a model of surface growth where atoms from an external source are deposited onto a surface and adsorbed atoms can evaporate. Depending on the nature of the incident flux, the mobility of the adatoms, and relaxation mechanisms, a wide variety of surface morphologies can arise. A surface is characterized by its height, $H(\mathbf{r}, t)$, as a function of transverse coordinates $\mathbf{r}$ and time t. It is more convenient to consider deviations of the surface from its average height, $h(\mathbf{r}, t) = H(\mathbf{r}, t) - \langle H \rangle$. A basic goal is to understand the properties of the height function h and its correlation function $\langle h(\mathbf{r}, t)h(\mathbf{r}', t') \rangle$.

Viewed as a surface evolution model, the Laplacian in (2.105) is positive near local surface minima (positive curvature) and negative near local maxima (negative curvature). Thus the Laplacian tends to smooth a surface and mimics the influence of the surface tension. The noise is assumed to be Gaussian with zero mean, independent of h, and spatially and temporally uncorrelated. That is,

$$\langle \eta \rangle = 0, \qquad \langle \eta(\mathbf{x}, t)\eta(\mathbf{x}', t') \rangle = 2\Gamma\delta(\mathbf{x} - \mathbf{x}')\,\delta(t - t'). \tag{2.106}$$

The competition between the effect of this noise and the Laplacian smoothing leads to a non-trivial surface morphology.

Let's compute the interface width $w(t) = \sqrt{\langle h^2(x,t)\rangle}$ for the one-dimensional situation with an initially flat interface, $h(x, t = 0) = 0$. As always, it is useful to start with dimensional analysis. By translational invariance, the width w is independent of x. Therefore $w = w(t, D, \Gamma)$. From (2.105) we find $[\eta] = L/T$, while (2.106) gives $[\Gamma] = [\eta]^2 LT = L^3/T$ (in one dimension). Thus dimensional analysis alone leads to a non-trivial scaling prediction

$$w^2 = Dt\, F(\kappa), \qquad \text{with} \quad \kappa = D^3 t/\Gamma^2, \tag{2.107}$$

for the width. We can do even better by exploiting the linearity of the EW equation (2.105). The height (and width) should be proportional to η, i.e. to $\sqrt{\Gamma}$, from (2.106). This dependence will certainly hold in the long-time limit, and is, in fact, valid for all times for an initially flat surface. Hence w is proportional to $\sqrt{\Gamma}$, implying that $F(\kappa) = C\kappa^{-1/2}$. Thus dimensional analysis and linearity determine the width

$$w^2 = C\Gamma\sqrt{t/D} \tag{2.108}$$

up to an amplitude C.

To derive (2.108) analytically and thereby compute the amplitude, we write the general solution of Eq. (2.105) in one dimension:

$$h(x,t) = \int_0^t dt_1 \int_{-\infty}^{\infty} dx_1\, \frac{\eta(x_1,t_1)}{\sqrt{4\pi D(t-t_1)}}\, e^{-(x-x_1)^2/4D(t-t_1)}\,. \tag{2.109}$$

That is, a noise input at (x_1, t_1) propagates to (x, t) via the Gaussian propagator of diffusion. Taking the square and averaging to obtain $w^2 = \langle h^2(x,t)\rangle$, we arrive at (2.108) with $C = \sqrt{2/\pi}$ (problem 2.28).

In two dimensions, dimensional analysis predicts

$$w^2 = Dt\, F(\kappa), \qquad \text{with} \quad \kappa = D^2 t/\Gamma\,. \tag{2.110}$$

Using this and the dependence $w \sim \sqrt{\Gamma}$ we arrive at the puzzling result that the width attains a stationary form: $w^2 = C\Gamma/D$. This stationarity is, in principle, feasible as the asymptotic behavior. However, owing to the lack of a characteristic time, the prediction $w^2 = C\Gamma/D$ (if correct!) would be valid at all times. This constancy is obviously impossible since we set the initial width equal to zero. (In fact, we shall see that the width is ill-defined, namely the computation of $w^2 = \langle h^2(\mathbf{x}, t)\rangle$ with $h(\mathbf{x}, t)$ given by (2.112), leads to $w = \infty$.)

To resolve this dilemma, let us look at the two-point correlation function

$$C(\mathbf{x}, t) = \langle h(\mathbf{x}, t)\, h(\mathbf{0}, t)\rangle\,. \tag{2.111}$$

This function is well-defined and finite for $\mathbf{x} \neq \mathbf{0}$. To compute $C(\mathbf{x}, t)$ we again use the general solution

$$h(\mathbf{x},t) = \int_0^t dt_1 \int d\mathbf{x}_1\, \frac{\eta(\mathbf{x}_1,t_1)}{4\pi D(t-t_1)}\, \exp\left[-\frac{(\mathbf{x}-\mathbf{x}_1)^2}{4D(t-t_1)}\right] \tag{2.112}$$

together with (2.106) to find (see problem 2.29)

$$C(\mathbf{x}, t) = \frac{\Gamma}{4\pi D} E_1(\xi), \qquad \xi = \frac{\mathbf{x}^2}{8Dt}, \tag{2.113}$$

where $E_1(z)$ is the exponential integral.[20] Using the asymptotic behavior

$$E_1(z) = -\ln z - \gamma_E + z - \frac{z^2}{4} + \frac{z^3}{18} + \cdots$$

($\gamma_E \doteq 0.577215$ is the Euler constant), we see that

$$C(\mathbf{x}, t) = \frac{\Gamma}{4\pi D} \left[\ln\left(\frac{8Dt}{\mathbf{x}^2}\right) - \gamma_E \right] \tag{2.114}$$

in the long-time limit.

The divergence of the width is resolved by noting that, in any physical situation, there is always a microscopic length a below which the assumption that noise is spatially uncorrelated is no longer valid. This cutoff is the natural shortest length beyond which the continuum description has no meaning. In lattice problems, a is the lattice spacing; in surface deposition problems, a is the atomic size. Mathematically, the factor $\delta(\mathbf{x} - \mathbf{x}')$ in the noise correlation function should be replaced by a function $f(\mathbf{x} - \mathbf{x}')$ that has the following properties:

- $f(\mathbf{0})$ is finite;
- $f(\mathbf{y})$ is a rapidly decaying function when $|\mathbf{y}| \gg a$;
- $\int d\mathbf{y} f(\mathbf{y}) = 1$, so that the magnitude of the noise is captured by parameter Γ.

The finite lower cutoff cures the divergence and leads to a width $w = \sqrt{(\Gamma/4\pi D) \ln(Dt/a^2)}$ that diverges logarithmically in the cutoff.

While the two-dimensional case is most natural in applications to surface evolution, a general lesson of statistical physics (and indeed of many other branches of physics) is to consider the behavior in general spatial dimensions. This generality can uncover regularities that may shed light on the behavior in the physically interesting dimension. The EW equation is tractable in any dimension, and the habit of considering general spatial dimensions is often useful. In the present case, the behavior of the two-point correlation function vividly demonstrates the importance of the spatial dimension:

$$C(\mathbf{x}, t) = \frac{\Gamma}{4\pi D} \times \begin{cases} (8\pi Dt)^{1/2} E_{3/2}(\xi), & d = 1, \\ E_1(\xi), & d = 2, \\ (8\pi Dt)^{-1/2} E_{1/2}(\xi), & d = 3. \end{cases} \tag{2.115}$$

In three dimensions, the two-point correlation function becomes time-independent in the long-time limit. From (2.115) we arrive at the following asymptotic behaviors:

$$w^2 = \Gamma \times \begin{cases} \sqrt{2t/\pi D}, & d = 1, \\ (4\pi D)^{-1} \ln(Dt/a^2), & d = 2, \\ (4\pi Da)^{-1}, & d = 3. \end{cases} \tag{2.116}$$

[20] The exponential integrals $E_n(z)$ are defined as $E_n(z) = \int_1^\infty du\, u^{-n} e^{-uz}$.

Thus noise is able to roughen an interface that evolves by the EW equation only when $d \leq 2$.

Although the width and more generally $C(x, t)$ remain non-stationary in one dimension, the height difference reaches a stationary limit. This feature can be appreciated by averaging the mean-square height. Using (2.115) we obtain

$$\langle [h(x, t) - h(0, t)]^2 \rangle = 2[C(0, t) - C(x, t)]$$

$$= \frac{\Gamma}{2\pi D} (8\pi D t)^{1/2} [E_{3/2}(0) - E_{3/2}(\xi)]. \qquad (2.117)$$

Writing

$$E_{3/2}(0) - E_{3/2}(\xi) = \int_1^\infty \frac{du}{u^{3/2}} \left[1 - e^{-u\xi} \right] \qquad (2.118)$$

and differentiating with respect to ξ we find that the derivative approaches to $\sqrt{\pi/\xi}$ as $\xi \to 0$. Therefore the difference (2.118) behaves as $2\sqrt{\pi\xi}$ when $\xi \to 0$. Substituting this result into (2.117) gives

$$\langle [h(x, t) - h(0, t)]^2 \rangle = \frac{\Gamma}{D} |x| \quad \text{as} \quad t \to \infty. \qquad (2.119)$$

This neat formula tells us that, in one dimension, the height $h(t, x)$ is (asymptotically) a random walk with respect to the spatial coordinate x.

The EW equation gives a flavor of the vast subject of Langevin equations that are built upon partial differential equations. These equations are rarely solvable; those which are tend to be linear. One such solvable equation is the Mullins equation (see problem 2.31)

$$\frac{\partial h}{\partial t} = -\nu \nabla^4 h + \eta. \qquad (2.120)$$

Similar (but often nonlinear) equations describe the phase-ordering kinetics at a positive temperature.[21]

Kardar–Parisi–Zhang equation

Similar to the EW equation, the Kardar–Parisi–Zhang (KPZ) equation has been proposed to mimic surface growth. This leads to a nonlinear Langevin equation that is much more challenging than the EW equation. Both the KPZ and EW equations assume that the surface is characterized by its height $h(\mathbf{x}, t)$, and thereby $h(\mathbf{x}, t)$ is considered as a single-valued function; physically, this means that there are no overhangs in the surface profile. Further, both equations are examples of a macroscopic description, where we are interested in spatial scales that greatly exceed the atomic scale. The chief distinction between the EW and KPZ

---

[21] More about this topic appears in Chapter 9.

equations is that the KPZ equation is nonlinear. The KPZ equation has the simple form[22]

$$\frac{\partial h}{\partial t} = D\nabla^2 h + \lambda(\nabla h)^2 + \eta, \tag{2.121}$$

yet the simplicity is illusory – the properties of this equation have been understood only in one spatial dimension.

It is worth emphasizing that Eq. (2.121) is not a random example from a zoo of non-linear Langevin equations – it is a unique specimen. Indeed, the *simplest* equation that is compatible with the symmetries of the problem is usually the most appropriate. From this perspective, notice that (2.121) is consistent with the following symmetries:

1. *Translation invariance along the growth direction.* This disallows the appearance of terms that involve the height function h explicitly; only derivatives of h can appear.
2. *Translation, inversion, and rotation symmetries in the basal plane* **x**. These symmetries permit terms like $\nabla^2 h$, $\nabla^4 h$, even powers of gradient $(\nabla h)^2$, $(\nabla h)^4$, etc., but exclude odd powers of gradient (terms such as ∇h).

In principle, an equation of the form

$$\frac{\partial h}{\partial t} = D_1\nabla^2 h + D_2\nabla^4 h + \lambda_1(\nabla h)^2 + \lambda_2(\nabla h)^4 + \mu(\nabla^2 h)(\nabla h)^2 + \eta$$

is compatible with the above symmetries and it may even provide a better early-time description of a particular process than (2.121). Asymptotically, however, higher-order derivatives and higher-order terms are negligible, that is, $\nabla^2 h \gg \nabla^4 h$, $(\nabla h)^2 \gg (\nabla h)^4$, $(\nabla h)^2 \gg (\nabla^2 h)(\nabla h)^2$, and therefore we recover (2.121). If the process is additionally invariant under up/down symmetry, $h \to -h$, the nonlinear terms should vanish $(\lambda_1 = \lambda_2 = 0)$ and we recover (2.105); if a symmetry additionally implies $D_1 = 0$, we obtain (2.120). These arguments explain the privileged role played by special equations, such as the EW equation, the Mullins equation, and the KPZ equation.

Note that the EW equation enjoys up/down symmetry, that is, symmetry under the transformation $h \longrightarrow -h$, which does not hold for the KPZ equation. This disparity suggests that if surface growth proceeds via deposition and evaporation and both processes balance each other, then the EW equation should provide a faithful description. If, on the other hand, one of the processes dominates, up/down symmetry is broken[23] and the EW equation (and more generally any Langevin equation that is linear in h) is inapplicable.

Let's try to extract as much as we can from the KPZ equation in the laziest way, that is, by dimensional analysis. For concreteness, we focus on the one-dimensional system. The width of the interface is a function of time and the parameters D, λ, Γ of the KPZ equation: $w = w(t, D, \lambda, \Gamma)$. Writing the dimensions of these variables, $[t] = T$, $[D] = L^2/T$, $[\lambda] = L/T$, and $[\Gamma] = L^3/T$, we see that two dimensionless complexes can be formed out of the

---

[22] Align (2.121) is a Langevin extension of the Burgers equation (with a somewhat unusual noise term). Indeed, applying the gradient operator ∇ to (2.121) we find that $\mathbf{u} = -2\lambda\nabla h$ satisfies $\mathbf{u}_t + (\mathbf{u} \cdot \nabla)\mathbf{u} = D\nabla^2\mathbf{u} + \boldsymbol{\zeta}$, which is the Burgers equation with noise $\boldsymbol{\zeta} = -2\lambda\nabla\eta$.

[23] Even in the reference frame associated with the average height.

four variables. Thus dimensional analysis gives

$$w = \sqrt{Dt}\, F(D^3 t/\Gamma^2, \Gamma\lambda/D^2)\,. \tag{2.122}$$

Recall that in the case of the EW equation we expressed the width in terms of one dimensionless variable $\kappa = D^3 t/\Gamma^2$ and then exploited the linearity of the EW equation to determine the width. For the nonlinear KPZ equation, the width is expressed as an unknown function of two variables and no further simplification appears feasible.

We now describe a useful trick that is surprisingly effective in the present context. The idea is to artificially *increase* the number of dimensions. In surface growth problems, there is an obvious difference between the lateral and vertical directions, and this suggests postulating different dimensions corresponding to these directions. We write $L = [x]$ for the transverse and $H = [h]$ for the vertical dimensions. Using (2.106) we get $[\Gamma] = [\eta]^2 LT$, while (2.121) gives $[\eta] = H/T$ and $[\lambda] = L^2/HT$. Therefore the dimensions of the relevant variables are

$$[t] = T, \quad [D] = \frac{L^2}{T}, \quad [\lambda] = \frac{L^2}{HT}, \quad [\Gamma] = \frac{H^2 L}{T}\,. \tag{2.123}$$

We want to find how the width w and the correlation length ℓ (which is the typical lateral separation between the points when their heights become uncorrelated) depend on the above four variables. We can form only one dimensionless variable $\Gamma^2\lambda^4 t/D^5$ out of the four variables (2.123). Taking into account that $[w] = H$ and $[\ell] = L$ and using dimensional analysis we obtain

$$w = \frac{D}{\lambda} F(\tau), \quad \ell = \sqrt{Dt}\, G(\tau), \quad \tau = \frac{\Gamma^2\lambda^4 t}{D^5}\,. \tag{2.124}$$

It is striking that a simple formal trick has brought such a huge simplification. One may wonder about the validity of the trick in the case of the EW equation. For the EW equation, we have three variables t, D, and Γ with independent dimensions, so dimensional analysis gives the full dependence on the parameters:

$$w \sim \Gamma^{1/2} (t/D)^{1/4}, \quad \ell \sim \sqrt{Dt}\,.$$

Thus for the EW equation, the application of this trick allows us to deduce the basic dependence of the width on physical parameters without invoking the linearity of the EW equation.

Unfortunately dimensional analysis alone for the KPZ equation cannot do better than give the scaling predictions of Eq. (2.124), which still involve unknown single-variable functions. It is also unknown how to proceed in two and higher dimensions. The one-dimensional case, however, is unique because a stationary solution to the Fokker–Planck equation for the probability distribution of h is known. The Fokker–Planck equation is the master equation for the probability distribution $\Pi[h]$. For the KPZ equation (and generally for the Fokker–Planck description of any partial differential equation), we must take into account that the probability distribution is a *functional* $\Pi[h(\mathbf{x})]$. Therefore partial derivatives with respect to h that appear in the Fokker–Planck equation should be replaced by functional derivatives.

In the case of the KPZ equation, the standard recipe of deriving the Fokker–Planck from the Langevin equation gives (see the discussion leading to Eq. (2.104))

$$\frac{\partial \Pi}{\partial t} = - \int d\mathbf{x} \, \frac{\delta}{\delta h} \left[\left(D\nabla^2 h + \lambda (\nabla h)^2 \right) \Pi \right] + \Gamma \int d\mathbf{x} \, \frac{\delta^2 \Pi}{\delta h^2} \, . \qquad (2.125)$$

In one dimension, this equation admits the stationary solution

$$\Pi = \exp \left[- \int dx \, \frac{D}{2\Gamma} \left(\frac{\partial h}{\partial x} \right)^2 \right] . \qquad (2.126)$$

This form does *not* provide a solution in higher dimensions $d \geq 2$.

The simplest way to determine the growth laws in the one-dimensional KPZ equation is to notice that the parameters D and Γ appear only in the combination Γ/D in the long-time limit. Using this key property, we employ dimensional analysis[24] to seek the width and the correlation length as functions of t, λ, and Γ/D. These parameters have independent dimensions, and therefore we arrive at the full asymptotic dependence on model parameters:

$$w \sim (\lambda t)^{1/3} (\Gamma/D)^{2/3}, \qquad \ell \sim (\lambda t)^{2/3} (\Gamma/D)^{1/3}. \qquad (2.127)$$

These scaling behaviors are consistent with (2.124) when $F \sim \tau^{1/3}$ and $G \sim \tau^{1/6}$. The $t^{1/3}$ growth of the width is often called the KPZ growth and the 1/3 exponent is called the KPZ exponent.[25]

2.10 Notes

The field of random walks, diffusion, and first-passage processes are central topics in probability theory and there is a correspondingly vast literature. In addition to the references mentioned in the previous chapter, we recommend [12, 13] for more probabilistic aspects of random walks and probability theory in general. Useful and comprehensive review articles about diffusion in random media are [14, 15]. The lack of self-averaging is reviewed in [15, 16]. The Holtsmark distribution and its applications to astronomy are given in Chandrasekhar's review article [12]. More comprehensive accounts of first-passage properties are given in [13, 17]. Various aspects of vicious random walks, and one-dimensional capture problems in particular, are described in [18–20].

An insightful discussion of the relation between the electrical capacitance of conducting objects and the reaction rate of these same objects appears in [21, 22]. The use of the quasi-static methods to solve diffusion equations in the presence of moving boundaries is discussed, for example, in [23–27]. The Langevin equation is extensively discussed in

---

[24] The extended dimensional analysis with two length units, $[w] = H$ and $[\ell] = L$, is required here.

[25] The exponent value of 1/3 first appeared in the work of Burgers on the deterministic Burgers equation with random initial conditions, long before the paper by Kardar, Parisi, and Zhang to which this exponent is commonly attributed. It took a long time to recognize the universality of this behavior.

Chandrasekhar's review article [12] and van Kampen's book [13]. The latter extensively discusses the connection between the Langevin equation and the corresponding Fokker–Planck equation. For applications of Langevin-type equations to surface growth phenomena, see [28–30].

2.11 Problems

2.1 A discrete random walk hops to the right with probability p and to the left with probability $q = 1 - p$ at each step. Let $P_N(x)$ be the probability that the particle is at x at the Nth time step.

(a) Write the master equation for this occupation probability.

(b) For the initial condition $P_0(x) = \delta_{x,0}$, show that the combined Fourier transform and generating function

$$P(k,z) = \sum_{N \geq 0} z^N \sum_{x=-\infty}^{\infty} e^{ikx} P_N(x)$$

is given by $P(k,z) = [1 - zu(k)]^{-1}$, where $u(k) = p\,e^{ik} + q\,e^{-ik}$ is the Fourier transform of the single-step hopping probability.

(c) Invert the Fourier transform and the generating function to determine the probability distribution of the discrete random walk.

2.2 Consider a discrete random walk in one dimension in which a step to the right of length 2 occurs with probability 1/3 and a step to the left of length 1 occurs with probability 2/3. Investigate the finite-time corrections to the asymptotic isotropic Gaussian probability distribution. *Hint:* Study the behavior of moments beyond second order, $\langle x^k \rangle$ with $k > 2$.

2.3 Solve Eq. (2.4) for the continuous-time random walk by first Laplace-transforming to give the difference equation

$$sP(n,s) - P(n, t = 0) = \frac{1}{2}P(n+1, s) + \frac{1}{2}P(n-1, s) - P(n, s),$$

for the Laplace transform $P(n,s) = \int_0^\infty P_n(t)\, e^{-st}\, dt$. Show that $P(n,s) = \lambda_-^n/(s + 1 - \lambda_-)$, where $\lambda_\pm = (1 \pm \sqrt{1 - 4a^2})/2a$ and $a = [2(s+1)]^{-1}$. Show that in the long-time limit, corresponding to $s \to 0$ in the Laplace transform, inversion of the Laplace transform recovers the Gaussian distribution (2.6) for $P_n(t)$.

2.4 Generalize Eq. (2.4) for the hypercubic lattice in d dimensions and find the probability for the walk to be at a given site. Verify that the probability for being at the origin is given by Eq. (2.46).

2.5 *Continuum limit of the master equation* (2.1).

(a) For the case of symmetric hopping, $p = q = \frac{1}{2}$, derive the diffusion equation (2.7) with $D = \frac{1}{2}$ by expanding the governing equation in a Taylor series to lowest non-vanishing order.

(b) Generalize the derivation to the case of non-symmetric hopping ($p \neq q$) and show that the continuum limit is the convection–diffusion equation

$$\frac{\partial P(x,t)}{\partial t} + v\frac{\partial P(x,t)}{\partial x} = D\frac{\partial^2 P(x,t)}{\partial x^2}.$$

Determine the diffusion coefficient D and the bias velocity v in terms of the microscopic parameters (p and q) of the random walk.

(c) For the initial condition $P(x, t = 0) = \delta(x)$, show that the solution to the convection–diffusion equation is

$$P(x,t) = \frac{1}{\sqrt{4\pi Dt}}\, e^{-(x-vt)^2/4Dt}.$$

2.6 Solve the convection–diffusion equation by the Laplace transform method and show that the Laplace transform of the probability distribution, $P(x,s)$, is

$$P_{\pm}(x,s) = \frac{1}{\sqrt{v^2 + 4Ds}}\, e^{-\alpha_{\mp}|x|},$$

where the subscript $\pm$ denotes the distribution in the regions $x > 0$ and $x < 0$, respectively, and $\alpha_{\pm} = \left(v \pm \sqrt{v^2 + 4Ds}\right)/2D$. Invert this Laplace transform to recover the Gaussian probability distribution.

2.7 Solve the convection–diffusion equation on the half-line $x > 0$, subject to the absorbing boundary condition $P(x = 0, t) = 0$ and the initial condition $P(x, t = 0) = \delta(x - \ell)$.

2.8 Suppose that N independent identically distributed random variables are drawn from the probability distribution $p(x)$. Estimate the largest of these variables for the cases: (a) $p(x) = e^{-x}$ for x positive; (b) $p(x) = \mu x^{-(1+\mu)}$ for $x > 1$ and $p(x) = 0$ for $0 < x < 1$; and (c) $p(x) = 1$ for $0 < x < 1$ and $p(x) = 0$ otherwise. As a corollary, estimate the speed of the most energetic molecule in a typical classroom.

2.9 Investigate a discrete-time random walk with a symmetric and broad step length distribution that has an algebraic tail: $p(x) \simeq A|x|^{-1-\mu}$ when $|x| \to \infty$. Assume that $0 < \mu < 1$.

(a) Write the Fourier transform of the step length distribution in the form

$$p(k) = 1 + \int_{-\infty}^{\infty} (e^{ikx} - 1)p(x)\, dx,$$

and show that the second term behaves as $-2A\Gamma(-\mu)|k|^{\mu}$ when $k \to 0$.

(b) Substitute $p(k) = 1 + A\Gamma(-\mu)|k|^{\mu}$ into (2.18) and show that the displacement distribution attains the scaling form

$$P_N(X) \to N^{-1/\mu} L_{\mu,0}(z), \quad z = X/N^{1/\mu},$$

with $L_{\mu,0}(z)$ given by (2.27) with $C = -2A\Gamma(-\mu)$.

2.10 Suppose that a discrete-time random walk has the step length distribution $p(x) = \pi^{-1}(1 + x^2)^{-1}$ (the Cauchy distribution).
  (a) Show that the corresponding Fourier transform is $p(k) = e^{-|k|}$.
  (b) Compute the integral in (2.18) when $p(k) = e^{-|k|}$ and show that

$$P_N(X) = \frac{1}{\pi} \frac{N}{N^2 + X^2},$$

  for any integer $N \geq 1$. Thus the Cauchy distribution is not merely a stable law in the $N \to \infty$ limit, it is stable for all N, a feature that characterizes the Gaussian distribution.

2.11 This problem concerns the derivation of the amplitude in (2.37).
  (a) Use $\mathbf{f}$ instead of $\mathbf{r}$ as an integration variable in (2.36). Show that $d\mathbf{r} = -\frac{1}{2}f^{-9/2}d\mathbf{f}$, so that (2.36) becomes

$$\Phi(\mathbf{k}) = \frac{1}{2}\int \frac{d\mathbf{f}}{f^{9/2}}\left(1 - e^{i\mathbf{k}\cdot\mathbf{f}}\right).$$

  (b) Show that the above integral is unaffected if we replace $\mathbf{f}$ by $-\mathbf{f}$. Taking the arithmetic mean of these two integrals show that

$$\Phi(\mathbf{k}) = \frac{1}{2}\int \frac{d\mathbf{f}}{f^{9/2}}\left[1 - \cos(\mathbf{k}\cdot\mathbf{f})\right].$$

  An alternative derivation is to notice that the integral in (a) must have a real value.
  (c) Introduce spherical coordinates in $\mathbf{f}$-space with z axis along $\mathbf{k}$, so that $d\mathbf{f} = 2\pi \sin\theta\, d\theta\, f^2 df$. Show that the integral from (b) becomes

$$\Phi(\mathbf{k}) = \pi \int_0^\infty \frac{df}{f^{5/2}} \int_{-1}^1 d\mu\, [1 - \cos(kf\mu)], \quad \mu = \cos\theta.$$

  (d) Perform the integration over μ and show that the integral from (c) simplifies to

$$\Phi(\mathbf{k}) = 2\pi k^{3/2} \int_0^\infty dz\, \frac{z - \sin z}{z^{7/2}}, \quad z = kf.$$

  (e) Compute the integral from part (d) and recover (2.37) with amplitude $a = \frac{4}{15}(2\pi)^{3/2}$.

2.12 Suppose that stars are distributed randomly (no correlations) and uniformly with density n. What is the average distance between a star and its closest neighbor?
  (a) On dimensional grounds show that the average distance scales as $n^{-1/3}$, or more generally as $n^{-1/d}$ in d dimensions.
  (b) Show that in d dimensions the average distance equals $\Gamma(1/d)\,(V_d n)^{-1/d}$, where V_d is the volume of a unit ball in d dimensions.
  (c) Show that $V_d = \pi^{d/2}/\Gamma(1 + d/2)$.
    (*Hint*: Work with the integral $J_d = \int e^{-r^2}\, d\mathbf{r}$ over all space $\mathbb{R}^d$. Compute this integral by two methods. First, write $r^2 = x_1^2 + \cdots + x_d^2, d\mathbf{r} = dx_1 \ldots dx_d$ to

recast J_d into the product, $J_d = (J_1)^d$, where $J_1 = \int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}$. Second, using rotational invariance and spherical coordinates, reduce the integral to $J_d = \int_0^{\infty} e^{-r^2} \Omega_d r^{d-1} dr$ and express this integral via the gamma function. Comparing the two answers will give the surface "area" Ω_d of the unit sphere; the volume of the unit ball is then determined from $\Omega_d = dV_d$.)

2.13 For $\mathbf{r} \neq 0$ and $t > 0$, explicitly verify Eq. (2.44) in one dimension.

2.14 Derive (2.54) from (2.53).

2.15 Explain the connection between the flux and the exit time probability distribution and thereby complete the derivation of (2.64).

2.16 Show that the solution to Eq. (2.72) on the interval $(0, L)$ is $t(x) = x(L - x)/(2D)$.

2.17 Generalize Eq. (2.72) by deriving the differential equations for the higher moments, $\langle T^2(x) \rangle$ and $\langle T^3(x) \rangle$. Solve these equations on the interval $(0, L)$.

2.18 Consider Eq. (2.77) on the interval $(-L, L)$.

(a) Show that the Laplace transform of the exit time is given by

$$\Pi(s, x) = \frac{\cosh[x\sqrt{s/D}]}{\cosh[L\sqrt{s/D}]}.$$

Suppose that a random walker starts at the origin, $x = 0$. Expand $\Pi(s, x = 0)$ in powers of s and deduce from this expansion that $\langle T^n \rangle = A_n \tau^n$, where $\tau = L^2/D$ and the coefficients are

$$A_1 = \frac{1}{2}, \quad A_2 = \frac{5}{24}, \quad A_3 = \frac{61}{720}, \quad A_4 = \frac{277}{8064},$$

$$A_5 = \frac{50521}{3628800}, \quad A_6 = \frac{540553}{95800320}, \quad \dots.$$

(b) Show that the general expression for the nth moment of the exit time is

$$\langle T^n \rangle = \frac{|E_{2n}|}{(2n)!} \tau^n,$$

where E_n are the Euler numbers that are defined by the expansion

$$\frac{1}{\cosh z} = \sum_{n \geq 0} \frac{E_n}{n!} z^n.$$

2.19 Suppose that a particle initially at $x = \ell$ undergoes unbiased diffusion with diffusion constant D. The particle is confined to the half-line (x_*, ∞), in which the boundary $x_*(t)$ is moving. Suppose that $x_* = Vt$, i.e. the boundary is initially at the origin and it approaches the particle with velocity $V > 0$. The exit time is the time when the particle first crosses the moving boundary.

(a) Show that the average exit time does not depend on the diffusion coefficient: $\langle T \rangle = \ell/V$.

(b) Show that the mean-square exit time is $\langle T^2 \rangle = (\ell/V)^2 + 2\ell D/V^3$.

(c) Show that the Laplace transform of the exit time $\langle e^{-sT} \rangle$ is given by $\exp[-\ell(\sqrt{V^2 + 4sD} - V)/(2D)]$.

Examine the same problem when the boundary is receding: $x_* = -Vt$.

(d) Show that the exit time is infinite with probability $\text{Prob}(T = \infty) = 1 - e^{-\ell V/D}$.

(e) Show that the Laplace transform of the exit time is $\exp[-\ell(\sqrt{V^2 + 4sD} + V)/(2D)]$.

2.20 A diffusing particle is confined to the interval $[0, L]$. The particle has a diffusivity D and it is also experiences a constant bias, with bias velocity v. The endpoint $x = L$ is reflecting, while if the particle reaches $x = 0$ it is adsorbed.

(a) Calculate the mean and the mean-square time for the particle to get trapped at $x = 0$ when it starts at $x = L$. Examine in detail the three cases of $v > 0$, $v < 0$, and $v = 0$.

(b) Compute the Laplace transform of the distribution of trapping times for the cases of $v > 0$, $v < 0$, and $v = 0$, and discuss the asymptotic behaviors of these distributions in the limits $t \to 0$ and $t \to \infty$.

2.21 A Brownian particle begins at the center of the three-dimensional ball of radius R.

(a) Show that the Laplace transform of the exit time is

$$\langle e^{-sT} \rangle = \frac{\sqrt{s\tau}}{\sinh\sqrt{s\tau}}, \quad \tau = R^2/D.$$

(b) Derive the general expression for the integer moments of the exit time

$$\langle T^n \rangle = (2^{2n} - 2) \frac{|B_{2n}|}{(2n)!} \tau^n$$

in terms of Bernoulli numbers B_m that are defined by the expansion

$$\frac{z}{e^z - 1} = \sum_{n \geq 0} \frac{B_n}{n!} z^n.$$

2.22 Consider the *gambler's ruin problem*. Two players A and B gamble repeatedly between themselves by betting \$1 in each trial. The probability of player A winning in a single bet is p. Assume that player A starts with n dollars and player B with $N - n$ dollars, and the game ends when one player has no money.

(a) What is the probability $\mathcal{E}_n$ that player A eventually wins?

(b) How long does the game last?

(c) Given that player A wins, how long does the game last? This setting defines the *conditional* exit time.

2.23 Solve the diffusion equation exterior to an absorbing sphere of radius a in one and two dimensions and thereby compute the exact flux to this sphere.[26] That is, solve

$$\frac{\partial c}{\partial t} = D\nabla^2 c \quad \text{with} \quad c(a = t, t) = 0, \quad c(r > a, t = 0) = 1,$$

---

[26] Here a must be greater than zero in $d = 2$ but can be set to zero for $d = 1$.

and then compute the time-dependent flux onto the sphere,

$$K(t) = D \int_{r=a} \nabla c(\mathbf{r}, t) \cdot d\boldsymbol{\sigma} .$$

2.24 A gas of spherical non-interacting particles of radii a are captured by a ball of radius R centered at the origin. In each capture event the volume of this ball increases by the volume $4\pi a^3/3$ of the captured particle.[27] Far from the absorbing ball, the concentration is c_∞.

(a) Using the stationary value of the reaction rate, $K = 4\pi DRc_\infty$, show that this capture process results in a "diffusive" growth of the radius $R(t)$ of the absorbing ball:

$$R = \sqrt{\mathcal{A}Dt}, \qquad \mathcal{A} = \frac{8\pi}{3} c_\infty a^3.$$

(b) Argue that the usage of the stationary reaction rate would be correct if the predicted growth was slower than diffusive. Justify the validity of the scaling ansatz

$$c(r,t) = c_\infty f(\eta), \qquad \text{where} \qquad \eta = \frac{r}{R} \quad \text{and} \quad R = \sqrt{\mathcal{A}Dt}$$

with as-yet unknown amplitude $\mathcal{A}$.

(c) Substitute the above scaling ansatz into (2.83) and show that the scaled concentration is

$$f(\eta) = 1 - \frac{g(\eta)}{g(1)}, \qquad \text{with} \qquad g(\eta) = \int_\eta^\infty \xi^{-2} e^{-\mathcal{A}\xi^2/4} \, d\xi.$$

(d) Use volume conservation to verify that the amplitude is an implicit solution of the equation

$$\frac{8\pi}{3} c_\infty a^3 = \mathcal{A} \, e^{\mathcal{A}/4} \int_1^\infty \xi^{-2} e^{-\mathcal{A}\xi^2/4} \, d\xi.$$

(e) In applications, the volume fraction of external particles is usually very low, $c_\infty a^3 \ll 1$, which implies that $\mathcal{A} \ll 1$. Show that in this situation the above relation for $\mathcal{A}$ approximately becomes

$$\frac{8\pi}{3} c_\infty a^3 = \mathcal{A} \left(\ln \frac{4}{\mathcal{A}} - \gamma_E \right),$$

where $\gamma_E \doteq 0.577215$ is the Euler constant.

2.25 Complete the derivation of Eqs (2.99) and (2.102).

2.26 A harmonically bound Brownian particle with non-zero mass obeys the equation of motion

$$\ddot{x}(t) + \gamma \dot{x}(t) + \omega_0^2 x(t) = \eta(t),$$

---

[27] This exercise is an example of a Stefan problem in which the boundary moves and its position $R = \sqrt{\mathcal{A}Dt}$ is determined in the process of solution.

where $\eta(t)$ is zero-mean white noise with $\langle \eta(t)\eta(t')\rangle = \Gamma\delta(t - t')$. For an arbitrary initial condition $x(0)$ and $\dot{x}(0)$, find $\langle x(t)\rangle$ and $\langle x(t)^2\rangle$. In the long-time limit $t \gg \gamma^{-1}$, show that your result is consistent with equilibrium statistical mechanics for a suitable choice of Γ.

2.27 Verify Eq. (2.113) for the two-point correlation function in two dimensions.

2.28 Fill in the missing steps to compute the exact form of mean-square width w^2 in Eq. (2.108) for the Edwards–Wilkinson model.

2.29 Confirm the result given in (2.115) for the two-point correlation function $C(x, t)$ in one dimension.

2.30 Analyze the Edwards–Wilkinson equation in three dimensions. Confirm the prediction of Eq. (2.115) for the two-point correlation function. Show that, when $x \ll \sqrt{Dt}$, the two-point correlation function simplifies to

$$C(\mathbf{x}, t) = \frac{\Gamma}{4\pi D}\frac{1}{x}.$$

2.31 Consider the Mullins equation (2.120).

(a) Use dimensional analysis and linearity to show that the width scales as $w \sim \Gamma^{1/2}\nu^{-d/8}t^{(4-d)/8}$. Therefore noise roughens the interface in the physically relevant case $d = 2$ and even in $d = 3$.

(b) Compute the width and the two-point correlation function in dimensions $d = 1$, 2, and 3.

2.32 Show that the probability distribution (2.126) is a solution of the Fokker–Planck equation (2.125) in one dimension.

2.33 Show that, for $d > 1$, the probability distribution (2.126) is not a solution to (2.125). (The integrand in Eq. (2.126) should contain $(\nabla h)^2$ in higher dimensions.)

3 Collisions

Non-equilibrium statistical physics courses usually begin with the Boltzmann equation and some of its most prominent consequences, particularly, the derivation of the Navier–Stokes equation of fluid mechanics and the determination of transport coefficients. Such derivations are cumbersome, often rely on uncontrolled approximations, and are treated in numerous standard texts. A basic understanding can already be gained by focusing on idealized collision processes whose underlying Boltzmann equations are sufficiently simple that they can be solved explicitly. These include the Lorentz gas, where a test particle interacts with a fixed scattering background, and Maxwell molecules, where the collision rate is independent of relative velocity. We also present applications of the Boltzmann equation approach to granular and traffic flows.

3.1 Kinetic theory

Non-equilibrium statistical physics originated in kinetic theory, which elucidated the dynamics of dilute gases and provided the starting point for treating more complex systems. Kinetic theory itself started with the Maxwell–Boltzmann velocity distribution, which was found before the Boltzmann equation – whose equilibrium solution is the Maxwell–Boltzmann distribution – had even been formulated.[1]

The Maxwell–Boltzmann distribution

Let's start by deriving the Maxwell–Boltzmann velocity distribution for a classical gas of identical molecules that is in equilibrium at temperature T. Two molecules scatter elastically when they are sufficiently close due to a short-range repulsive intermolecular potential. Let $P(\mathbf{v})\,d\mathbf{v}$ be the probability to find a molecule within a range $d\mathbf{v}$ about $\mathbf{v}$. The Maxwell–Boltzmann distribution can be derived using two fundamental assumptions:

1. *Rotational symmetry in velocity space*. This implies that the Maxwell–Boltzmann distribution is isotropic (i.e. it depends only on the speed $v = |\mathbf{v}|$ and not on the direction of $\mathbf{v}$):

$$P(\mathbf{v}) = P(v^2). \tag{3.1}$$

---

[1] James Clerk Maxwell derived this distribution in 1867, while the Boltzmann equation was first written in 1872.

2. *No correlations between different velocity components*. Together with symmetry, this assumption allows the velocity distribution to be factorized as

$$P(v^2) = p(v_x^2)\, p(v_y^2)\, p(v_z^2), \tag{3.2}$$

where $p(v_i^2)\, dv_i$ is the probability that the ith velocity component is in a range dv_i about v_i.

Using these assumptions, Maxwell deduced the Maxwell–Boltzmann distribution using a clever trick, namely taking the logarithm of (3.2)

$$\ln P(v^2) = \ln p(v_x^2) + \ln p(v_y^2) + \ln p(v_z^2)$$

and differentiating it with respect to velocity components. This gives

$$\frac{\partial \ln P(v^2)}{\partial v_i^2} = \frac{d \ln P}{dv^2}\frac{\partial v^2}{\partial v_i^2} = \frac{d \ln P}{dv^2} = \frac{d \ln p(v_i^2)}{dv_i^2},$$

and since this equation holds for any component, we conclude that the right-hand side must be constant. Therefore $d \ln P/dv^2 = -b$, from which $P = A\, e^{-bv^2}$. The constants in this velocity distribution are fixed by normalization, $\int P(\mathbf{v})\, d\mathbf{v} = n$, where n is the particle density, and the standard definition of temperature $\frac{1}{2}m\langle v_x^2 \rangle = \frac{1}{2}T$ to give

$$P(\mathbf{v}) = n \left(\frac{m}{2\pi T} \right)^{3/2} e^{-m\mathbf{v}^2/2T}. \tag{3.3}$$

Nothing about the interparticle potential entered into Maxwell's argument! We only need a scattering mechanism that conserves mass, momentum, and energy in each collision, so that a steady state actually exists. With these modest requirements, the steady-state velocity distribution is the Maxwell–Boltzmann form, independent of the intermolecular interaction, and also independent of the particle density.[2]

The Boltzmann equation

In classical kinetic theory, the fundamental quantity is the space- and velocity-dependent distribution, $P(\mathbf{r}, \mathbf{v}; t)\, d\mathbf{r}\, d\mathbf{v}$, defined as the number of particles within a range $d\mathbf{r}\, d\mathbf{v}$ about $\mathbf{r}, \mathbf{v}$ at time t. This distribution evolves in time according to

$$\left(\frac{\partial}{\partial t} + \mathbf{v} \cdot \frac{\partial}{\partial \mathbf{r}} + \mathbf{a} \cdot \frac{\partial}{\partial \mathbf{v}} \right) P(\mathbf{r}, \mathbf{v}; t)$$

$$= \int \sigma [P_2(\mathbf{r}, \mathbf{v}'; \mathbf{r}, \mathbf{w}') - P_2(\mathbf{r}, \mathbf{v}; \mathbf{r}, \mathbf{w})]$$

$$\times \delta(\mathbf{v} + \mathbf{w} - \mathbf{v}' - \mathbf{w}')\, \delta(v^2 + w^2 - v'^2 - w'^2)\, d\mathbf{w}\, d\mathbf{v}'\, d\mathbf{w}'. \tag{3.4}$$

---

[2] One may ask: Isn't it simpler still to "derive" by Maxwell–Boltzmann distribution from the Boltzmann factor, $e^{-E/T}$, of equilibrium statistical mechanics? While there is no work in writing the Boltzmann factor, considerable effort is involved in developing the apparatus of equilibrium statistical mechanics, upon which the Boltzmann factor is based. Maxwell's argument requires no apparatus at all!
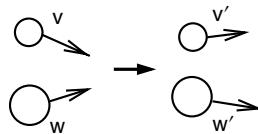
Fig. 3.1.   Binary collision, with pre-collision velocities (v, w) and post-collision velocities (v′, w′).

The terms on the left-hand side of Eq. (3.4) represent the change in $P(\mathbf{r}, \mathbf{v}; t)$ by particle motion in the absence of collisions; the last term is due to an external force that causes an acceleration $\mathbf{a}$. The right-hand side represents the change in $P(\mathbf{r}, \mathbf{v}; t)$ by collisions (Fig. 3.1). The loss term accounts for particles with phase space coordinates $(\mathbf{r}, \mathbf{v})$ that collide with particles with coordinates $(\mathbf{r}, \mathbf{w})$. The total collision rate involves the cross-section σ times the probability that the collision partners are in proximity,[3] as represented by the *two-body* correlation function P_2. The complementary gain term represents the rate at which particles with coordinates $(\mathbf{r}, \mathbf{v}')$ and $(\mathbf{r}, \mathbf{w}')$ collide, with one of the outgoing particles having coordinates $(\mathbf{r}, \mathbf{v})$ and the other $(\mathbf{r}, \mathbf{w})$. The delta functions in the integrand of Eq. (3.4) guarantee conservation of momentum and energy.[4]

Equation (3.4) is not closed because the single-particle distribution $P(\mathbf{r}, \mathbf{v}; t)$ is expressed in terms of a two-body distribution $P_2(\mathbf{r}, \mathbf{v}; \mathbf{r}, \mathbf{w})$. One therefore needs to write an equation for the two-body distribution that now involves the three-body distribution, etc. Continuing this construction gives an infinite hierarchy of equations that is known as the so-called BBGKY (Bogoliubov–Born–Green–Kirkwood–Yvon) hierarchy. Boltzmann suggested breaking this hierarchy at the very beginning by replacing the two-body distribution with a product of single-body distributions:

$$P_2(\mathbf{r}, \mathbf{v}; \mathbf{r}, \mathbf{w}) = P(\mathbf{r}, \mathbf{v}) P(\mathbf{r}, \mathbf{w}). \tag{3.5}$$

This replacement is based on the assumption that pre-collision velocities are uncorrelated – the *molecular chaos assumption* (originally coined by Boltzmann as the *Stosszahlansatz*). In many-particle systems, this lack of correlations is often referred to as *mean-field behavior*, and it allows one to express many-body distribution functions as a product of single-body distribution functions. With the molecular chaos assumption, we may now write a closed equation, the *Boltzmann equation* (BE), for the evolution of the particle distribution:

$$\left(\frac{\partial}{\partial t} + \mathbf{v} \cdot \frac{\partial}{\partial \mathbf{r}} + \mathbf{a} \cdot \frac{\partial}{\partial \mathbf{v}} \right) P(\mathbf{r}, \mathbf{v})$$

$$= \int \sigma \left[P(\mathbf{r}, \mathbf{v}') \, P(\mathbf{r}, \mathbf{w}') - P(\mathbf{r}, \mathbf{v}) \, P(\mathbf{r}, \mathbf{w}) \right]$$

$$\times \delta(\mathbf{v} + \mathbf{w} - \mathbf{v}' - \mathbf{w}') \, \delta(v^2 + w^2 - v'^2 - w'^2) \, d\mathbf{w} \, d\mathbf{v}' \, d\mathbf{w}' . \tag{3.6}$$

The molecular chaos assumption appears self-evident, yet it is subtle and it has been the source of much confusion. The most perplexing feature is how to reconcile microscopic

---

[3] Here we take proximity to mean the same spatial location.
[4] We assume that the molecules are identical; in particular, they have equal masses.

reversibility of the molecular motion with macroscopic *irreversibility*. Indeed, the under-lying microscopic (Newtonian) dynamics is reversible; that is, reversing the direction of time, $t \rightarrow -t$, the governing equations do not change. On the other hand, the world around us is irreversible on the macroscopic level. The molecular chaos assumption breaks the time-reversal symmetry and the resulting Boltzmann equation (3.6) describes observable irreversible behaviors. A classic example of this apparent paradox is to imagine that all molecules are initially in one half of a room. They will subsequently quickly spread over the entire room and the density will become essentially uniform. If, however, we reverse the velocities of all molecules, they should (at a certain moment) concentrate in a half of the room. To this objection Ludwig Boltzmann reportedly replied "go ahead, reverse them!" A more clear resolution, also suggested by Boltzmann, relies on probabilistic reasoning: simply put, irreversible behaviors are just much more plausible.

In addition to the molecular chaos assumption, it is important to keep in mind the other important assumptions underlying the BE and some of their implications:

1. The gas is dilute so that the one-body distribution $P(\mathbf{r}, \mathbf{v}; t)$ describes the state of the gas accurately. Furthermore, the gas is sufficiently dilute that only binary collisions are important.
2. Particles move with almost constant velocities between collisions and their trajectories bend only when particles are in a close proximity. A useful mental image is to view molecules as *hard spheres*.

While the conceptual problems associated with the BE are settled and the necessity to use statistical interpretation is no longer puzzling, the understanding of the BE is greatly hindered by its lack of analytical solutions. We understand theory by applying it to con-crete problems; for example, finding the energy levels of the harmonic oscillator or the hydrogen atom are crucial in developing an understanding of quantum mechanics. There is no such logical development in kinetic theory because of its mathematical complex-ity. The Boltzmann equation (3.6) is formally a nonlinear integro-differential equation in seven variables.[5] Moreover, for a realistic system, such as air that is composed mostly of diatomic molecules, accounting for vibrational and rotational degrees of freedom leads to many additional complexities.

Even when we attempt to simplify the BE as much as possible, we are faced with intractable problems. For example, one great success of the BE approach is that it led to the derivation of hydrodynamic equations, particularly the Navier–Stokes equations through the Chapman–Enskog expansion. This procedure is complicated and cumbersome and it has been fully implemented only for monatomic gases. Nevertheless, transport coefficients have not been computed even for the simplest case of the hard-sphere gas. Another route to simplification is to consider systems where spatial variation occurs in a single coordi-nate, as in the case of a shock wave. However, the microscopic particle velocities are still three-dimensional and one must account for all velocity components in solving the prob-lem, a complication that obviates the apparent simplification afforded by the variation of hydrodynamic variables along a single coordinate.

---

[5] In d dimensions, the one-body distribution $P(\mathbf{r}, \mathbf{v}; t)$ depends on $d + d + 1 = 2d + 1$ variables.

A still more drastic approach is to consider truly one-dimensional systems. However, the BE in one dimension is trivial – colliding equal-mass particles merely exchange their velocities. By relabeling the particles, we can view this exchange as two particles passing through each other with no interaction. Consequently, the velocity distribution does not evolve and there is no equilibrium state. To remedy the absence of interactions, we can consider a gas with distributed particle masses. However, the BE for such a mixture is again complicated. Moreover the validity of this BE is questionable because a particle can directly interact only with its left and right neighbors and the molecular chaos assumption becomes dubious.

The BE has a host of additional complicating features that are not completely resolved. For example, we have to supplement the BE with boundary conditions, and these are subtle for a situation as simple as a gas in a container. There are no "canonical" boundary conditions, and we must formally also consider the molecules in the walls and their interactions with the impinging gas molecules to define the boundary condition from first principles. It is ironic, but not surprising, that the BE has eluded any exact treatment. Because of this plethora of complications, we now turn to idealized descriptions of kinetic theory that avoid the technicalities in dealing with the BE. While these idealizations contain drastic over-simplifications to render them soluble, they are still grounded in underlying phenomenology and provide useful insights about physical kinetics.

3.2 The Lorentz gas

At the turn of the twentieth century, Paul Drude developed a qualitative theory for electrical conduction in metals. To establish a more solid basis for the Drude theory, Hendrik Lorentz suggested an idealized model for this electron transport in which: (i) electron–electron interactions are ignored, (ii) the background atoms are treated as immobile spherical scatterers, and (iii) the electron–atom interaction is described by elastic scattering. Assumption (ii) is crucial as it converts a many-body problem into an effective single-particle system that consists of a "test" particle and a static background. We now understand that quantum effects should be accounted for in this application to electrical conduction; however in many examples, e.g. in transport problems involving slow neutrons or in gas mixtures in which the mass ratio is large, the classical Lorentz model provides a good approximation. The Lorentz model represents a rare source of exact results in kinetic theory, it provides a concrete example where microscopic reversibility can be reconciled with macroscopic irreversibility, and it reveals a variety of unexpected properties.

The model

Consider a three-dimensional Lorentz model in which non-interacting classical particles (electrons) move among randomly distributed immobile hard spheres (atoms) of radii a. The electrons do not interact with each other and are elastically scattered by the hard spheres.

Consequently, only the direction of the electron velocity $\mathbf{v}$, and not its magnitude, changes in a collision. The probability density $f(\mathbf{r}, \mathbf{v}, t)$ to find an electron at position $\mathbf{r}$ with velocity $\mathbf{v}$ at time t evolves according to the Boltzmann equation

$$\left(\frac{\partial}{\partial t} + \mathbf{v} \cdot \frac{\partial}{\partial \mathbf{r}}\right) f(\mathbf{r}, \mathbf{v}, t) = \frac{1}{\tau} \left[\left(\int \frac{d\widehat{\mathbf{v}}}{4\pi} f(\mathbf{r}, \mathbf{v}, t)\right) - f(\mathbf{r}, \mathbf{v}, t)\right]. \tag{3.7}$$

Here $1/\tau = v/\ell$ is the collision frequency, $\ell = (n\pi a^2)^{-1}$ is the mean free path, and $\widehat{\mathbf{v}} = \mathbf{v}/v$. The left-hand side consists of the standard streaming terms of the Boltzmann equation. In the absence of collisions, the solution of Eq. (3.7) is $f(\mathbf{r}, \mathbf{v}, t) = f(\mathbf{r} - \mathbf{v}t, \mathbf{v}, 0)$, which describes a particle that moves at constant velocity. The right-hand side of Eq. (3.7) accounts for collisions. The loss term is proportional to the distribution multiplied by the collision frequency, while the gain term reflects the well-known feature that hard-sphere scattering in three dimensions is isotropic (problem 3.3). For notational convenience, we write the gain term as

$$[\mathbb{P}f](\mathbf{r}, \mathbf{v}, t) \equiv \int \frac{d\widehat{\mathbf{v}}}{4\pi} f(\mathbf{r}, \mathbf{v}, t), \tag{3.8}$$

with $\mathbb{P}$ the projection operator that averages over all directions of the velocity.

Equation (3.7) involves a number of important features that deserve emphasis:

1. Equation (3.7) is *linear*. This represents a huge simplification compared to the nonlinear Boltzmann equation. To stress this difference, Eq. (3.7) is often called the *Boltzmann–Lorentz* (BL) equation. Physically, linearity is a consequence of the scatterers being unaffected by collisions. Because of linearity, Eq. (3.7) is solvable, in contrast to the general intractability of the Boltzmann equation.
2. The simplicity of the collisional gain term, $[\mathbb{P}f](\mathbf{r}, \mathbf{v}, t)$, is a special feature of three dimensions. In two dimensions, elastic hard-sphere scattering is *not* isotropic (see problem 3.3)!
3. It involves the dilute limit. That is, the mean free path $\ell \gg a$, or equivalently the volume fraction occupied by the scatterers is small, $na^3 \ll 1$.
4. Equation (3.7) also requires randomly located scatterers.[6]

The main outcome from Eq. (3.7) is that electrons diffuse in the long-time limit. That the diffusion constant should scale as $v\ell$ follows from dimensional analysis. From the solution to Eq. (3.7) that will be presented below, we shall learn that the precise answer is

$$D = \frac{1}{3} v\ell. \tag{3.9}$$

---

[6] In physical applications, such as electrons moving in crystals, scatterers are arrayed on a periodic lattice. The BL equation is invalid for such a geometry because unbounded free-particle trajectories can exist – an infinite "horizon." These trajectories lead to the mean-square displacement growing slightly faster than linearly, namely, $\langle r^2 \rangle \sim t \ln t$. This result has been proved in two dimensions, and is believed also to be correct in three dimensions.

The diffusion constant

Let us first solve for the distribution of velocities

$$F(\mathbf{v}, t) = \int d\mathbf{r} f(\mathbf{r}, \mathbf{v}, t). \tag{3.10}$$

By integrating (3.7) over the spatial coordinates, we find that $F(\mathbf{v}, t)$ satisfies

$$\frac{\partial}{\partial t} F(\mathbf{v}, t) = \frac{1}{\tau} \left\{ [\mathbb{P}F](v, t) - F(\mathbf{v}, t) \right\}. \tag{3.11}$$

We apply the projector operator $\mathbb{P}$ to (3.11) and use the fundamental property (often taken as the definition of projector operators), $\mathbb{P}^2 = \mathbb{P}$, to give

$$\frac{\partial}{\partial t} [\mathbb{P}F](v, t) = 0. \tag{3.12}$$

Thus the isotropic part of the velocity distribution $[\mathbb{P}F](v, t)$, which we denote as $\Phi(v)$, does not change. The solution of Eq. (3.11) is therefore

$$F(\mathbf{v}, t) = \Phi(v) + [F_0(\mathbf{v}) - \Phi(v)]e^{-t/\tau}, \tag{3.13}$$

so that the velocity distribution becomes isotropic exponentially quickly, with relaxation time τ.

For the initial condition $F_0(\mathbf{v}) = \delta(\mathbf{v} - \mathbf{u})$ the general solution (3.13) becomes

$$F(\mathbf{v}, t \,|\mathbf{u}, 0) = \frac{\delta(v - u)}{4\pi u^2} + \left[\delta(\mathbf{v} - \mathbf{u}) - \frac{\delta(v - u)}{4\pi u^2} \right] e^{-t/\tau}. \tag{3.14}$$

Using (3.14) we can compute the diffusion constant using the remarkable Green–Kubo formula[7] that expresses the diffusion constant via the velocity autocorrelation function. This formula states that[8]

$$D = \frac{1}{3} \int_0^\infty dt \, \langle \mathbf{v}(t) \cdot \mathbf{v}(0) \rangle_{\text{equil}}, \tag{3.15}$$

where the equilibrium velocity autocorrelation function is

$$\langle \mathbf{v} \cdot \mathbf{u} \rangle_{\text{equil}} = \int d\mathbf{v} \int d\mathbf{u} \, (\mathbf{v} \cdot \mathbf{u}) \, F(\mathbf{v}, t \,|\mathbf{u}, 0) \, \Phi_{\text{equil}}(u). \tag{3.16}$$

The velocity autocorrelation function involves the average over the isotropic equilibrium state $\Phi_{\text{equil}}(u) = \delta(u - v_0)/(4\pi v_0^2)$. Since the speed remains constant, we have

[7] Generally, the Green–Kubo formula, derived in the 1950s, relates transport coefficients, such as the thermal and electrical conductivities, and the shear and bulk viscosities, to appropriate correlation functions. The "Green–Kubo" formula for the diffusion constant goes back to Einstein, where it appeared in one of his papers published during his *annus mirabilis* of 1905.

[8] The numerical factor in front of the integral in Eq. (3.15) is specific to three dimensions; in d dimensions, the prefactor is d^{-1}. The derivation of the Einstein–Green–Kubo formula (3.15) is the subject of problem 3.4.

$|\mathbf{v}| = |\mathbf{u}| = v_0$, and the integrations in (3.16) are projector operators. The isotropic terms in $F(\mathbf{v}, t\,|\mathbf{u}, 0)$ in Eq. (3.14) are proportional to $[\mathbb{P}\mathbf{v}] \cdot [\mathbb{P}\mathbf{u}]$ and thus vanish, since $\mathbb{P}\mathbf{v} \equiv 0$. Only the anisotropic term that involves the factor $\delta(\mathbf{v} - \mathbf{u})$ provides a non-vanishing contribution. Keeping only this term, we have

$$\langle \mathbf{v} \cdot \mathbf{u} \rangle_{\text{equil}} = v_0^2\, e^{-t/\tau} \int \frac{d\widehat{\mathbf{v}}}{4\pi} = v_0^2\, e^{-t/\tau} \,. \tag{3.17}$$

Substituting this result into (3.15) and performing the integral gives $D = \frac{1}{3} v_0^2 \tau$, which is equivalent to (3.9).

Exact probability density

Having determined the velocity distribution, let's try to solve the full BL equation (3.7) for the probability density $f(\mathbf{r}, \mathbf{v}, t)$. The linearity of the BL equation suggests using the Laplace transform in time, as we are interested in $t \geq 0$, and the Fourier transform in space, as we consider the entire three-dimensional space $\mathbb{R}^3$. Introducing the Fourier–Laplace transform

$$f(\mathbf{k}, \mathbf{v}, s) = \int_0^\infty dt\, e^{-st} \int d\mathbf{r}\, e^{-i\mathbf{k}\cdot\mathbf{r}} f(\mathbf{r}, \mathbf{v}, t) \tag{3.18}$$

reduces Eq. (3.7) to

$$[s + i\mathbf{k} \cdot \mathbf{v}] f(\mathbf{k}, \mathbf{v}, s) = \frac{1}{\tau}(\mathbb{P} - 1) f(\mathbf{k}, \mathbf{v}, s) + f_0(\mathbf{k}, \mathbf{v}) \,, \tag{3.19}$$

where $f_0(\mathbf{k}, \mathbf{v}) = f(\mathbf{k}, \mathbf{v}, t = 0)$ is the Fourier transform of the initial probability density. After straightforward steps we obtain

$$f(\mathbf{k}, \mathbf{v}, s) = \frac{\tau^{-1}}{\tau^{-1} + s + i\mathbf{k} \cdot \mathbf{v}} [\mathbb{P}f](\mathbf{k}, \mathbf{v}, s) + \phi_0(\mathbf{k}, \mathbf{v}, s) \,, \tag{3.20}$$

where we use the shorthand notation

$$\phi_0(\mathbf{k}, \mathbf{v}, s) = \frac{f_0(\mathbf{k}, \mathbf{v})}{\tau^{-1} + s + i\mathbf{k} \cdot \mathbf{v}} \,. \tag{3.21}$$

To complete the solution, we need $\mathbb{P}f$. We therefore apply the projection operator $\mathbb{P}$ on Eq. (3.20). Since $\mathbb{P}f$ is already isotropic, the action of $\mathbb{P}$ on the first term on the right-hand side of (3.20) requires that we compute

$$\mathbb{P}\left[\frac{\tau^{-1}}{\tau^{-1} + s + i\mathbf{k} \cdot \mathbf{v}}\right] = \frac{1}{4\pi} \int_0^{2\pi} d\phi \int_{-1}^{1} d(\cos\theta) \frac{\tau^{-1}}{\tau^{-1} + s + ikv\cos\theta}$$

$$= \frac{1}{kv\tau} \tan^{-1} \frac{kv\tau}{s\tau + 1} \,.$$

Therefore the action of $\mathbb{P}$ on Eq. (3.20) yields

$$\mathbb{P}f = \left[\frac{1}{kv\tau} \tan^{-1} \frac{kv\tau}{s\tau + 1}\right] \mathbb{P}f + \mathbb{P}\phi_0 \,,$$

from which we may determine $\mathbb{P}f$. Substituting the result for $\mathbb{P}f$ into Eq. (3.20), we arrive at the full solution for the Fourier–Laplace transform that is valid for an *arbitrary* initial condition:

$$f(\mathbf{k}, \mathbf{v}, s) = \frac{\tau^{-1}}{\tau^{-1} + s + i\mathbf{k} \cdot \mathbf{v}} \left[1 - \frac{1}{kv\tau} \tan^{-1} \frac{kv\tau}{s\tau + 1}\right]^{-1} \mathbb{P}\phi_0 + \phi_0 \,. \qquad (3.22)$$

Inverting this exact solution to obtain $f(\mathbf{r}, \mathbf{v}, t)$ is hard even for the isotropic initial condition, such as

$$f(\mathbf{r}, \mathbf{v}, t = 0) = \delta(\mathbf{r}) \frac{\delta(v - v_0)}{4\pi v_0^2} \,.$$

Nevertheless, the Fourier–Laplace transform allows one to extract the diffusion coefficient, as we now discuss.

Alternative derivation of the diffusion constant

Let us now extract the diffusion constant from the exact solution (3.22). In the hydrodynamic regime of large spatial and temporal scales, the density

$$\rho(\mathbf{r}, t) = \int d\mathbf{v} f(\mathbf{r}, \mathbf{v}, t) \sim \mathbb{P}f \qquad (3.23)$$

should evolve according to the diffusion equation

$$\frac{\partial \rho(\mathbf{r}, t)}{\partial t} = D\nabla^2 \rho(\mathbf{r}, t) \,. \qquad (3.24)$$

The Fourier transform of the density therefore satisfies

$$\frac{\partial \rho(\mathbf{k}, t)}{\partial t} = -Dk^2 \rho(\mathbf{k}, t) \,, \qquad (3.25)$$

from which $\rho(\mathbf{k}, t) = e^{-Dk^2 t}$, since $\rho(\mathbf{k} = 0, t) = \int d\mathbf{r}\, \rho(\mathbf{r}, t) = 1$. Laplace-transforming $\rho(\mathbf{k}, t)$ yields

$$\rho(\mathbf{k}, s) = (Dk^2 + s)^{-1} \,. \qquad (3.26)$$

To extract the generic form (3.26) from (3.22), we note that in the hydrodynamic regime $t \gg \tau$. Since $s \sim t^{-1}$ and $k \sim t^{-1/2}$, we have $\tau^{-1} \gg s,\ kv$. Hence (3.22) becomes

$$f = \left[1 - \frac{1}{kv\tau} \tan^{-1} \frac{kv\tau}{s\tau + 1}\right]^{-1} \mathbb{P}\phi_0 + \phi_0 \,.$$

We expand the term in the square brackets by using $\tan^{-1} x = x - x^3/3 + \cdots$ to give

$$1 - \frac{1}{kv\tau} \tan^{-1} \frac{kv\tau}{s\tau + 1} = 1 - \frac{1}{1 + s\tau} + \frac{1}{3} \frac{(kv\tau)^2}{(1 + s\tau)^3} + \cdots = \tau s + \frac{1}{3} k^2 v^2 \tau^2 + \cdots,$$

from which we conclude that $D = v^2\tau/3$. Furthermore, in the hydrodynamic regime Eq. (3.21) reduces to $\phi_0 = \tau f_0$. With these simplifications, Eq. (3.22) asymptotically simplifies to

$$f = \frac{1}{Dk^2 + s}\, \mathbb{P}f_0\,,$$

from which we indeed obtain (3.26). This computation additionally gives the diffusion constant $D = v^2\tau/3$, as already derived via the Einstein–Green–Kubo formula.

Lorentz gas in one dimension

The BL equation is tractable in three dimensions, in part, because the isotropy of hard-sphere elastic scattering led to the simple form of the collisional gain term in the BL equation (3.7). In contrast, hard-sphere scattering is not isotropic when $d \neq 3$; backward scattering is favored in two dimensions, and in one dimension there is only backward scattering. Thus we cannot literally define the Lorentz gas in one dimension, as the particle will be caged between two adjacent scatterers. However, a Lorentz gas in one dimension can be defined by postulating that a particles passes through a scatterer with probability $1 - p$ and that elastic backscattering occurs with probability p. Naively, this process looks like a one-dimensional random walk and we will learn the extent to which this notion is correct.

The BL equation for this one-dimensional Lorentz gas is then

$$\left(\frac{\partial}{\partial t} + v\frac{\partial}{\partial x}\right)f(x,v,t) = \frac{f(x,-v,t) - f(x,v,t)}{\tau}\,, \tag{3.27}$$

where $pv/\ell = 1/\tau$. Because the velocity can only change its sign in a collision, the probability distribution is a combination of two distributions:

$$f(x,v,t) = \rho_+(x,t)\,\delta(v - v_0) + \rho_-(x,t)\,\delta(v + v_0)\,.$$

The BL equation (3.27) then splits into two equations that describe the evolution of the densities of the left- and right-moving particles:

$$\left(\frac{\partial}{\partial t} + v_0\frac{\partial}{\partial x}\right)\rho_+ = \frac{\rho_- - \rho_+}{\tau}, \tag{3.28a}$$

$$\left(\frac{\partial}{\partial t} - v_0\frac{\partial}{\partial x}\right)\rho_- = \frac{\rho_+ - \rho_-}{\tau}\,. \tag{3.28b}$$

We now recast these equations into a hydrodynamic form by replacing the densities $\rho_\pm(x,t)$ by the total density and the current, respectively,

$$\rho(x,t) = \rho_+(x,t) + \rho_-(x,t),$$
$$J(x,t) = v_0\left[\rho_+(x,t) - \rho_-(x,t)\right]\,.$$

Adding and subtracting Eqs (3.28a)–(3.28b) gives

$$\frac{\partial \rho}{\partial t} + \frac{\partial J}{\partial x} = 0, \tag{3.29a}$$

$$\frac{\partial J}{\partial t} + v_0^2 \frac{\partial \rho}{\partial x} = -\frac{2J}{\tau} . \tag{3.29b}$$

The first equation expresses mass conservation and could have been written without any derivation. The compactness of the second equation stems from the simplicity of the one-dimensional Lorentz model.

We now eliminate J by differentiating Eq. (3.29a) with respect to t and Eq. (3.29b) with respect to x, and then subtracting to obtain

$$\frac{\tau}{2} \frac{\partial^2 \rho}{\partial t^2} + \frac{\partial \rho}{\partial t} = \frac{v_0^2 \tau}{2} \frac{\partial^2 \rho}{\partial x^2} , \tag{3.30}$$

which is known as the *telegraph equation*. A few important lessons follow directly from Eq. (3.30):

1. The value of the diffusion constant is

$$D = \frac{1}{2} v_0^2 \tau . \tag{3.31}$$

2. The telegraph equation is *hyperbolic* so that perturbations propagate at a finite speed. In contrast, all standard hydrodynamic equations (diffusion, Navier–Stokes, etc.) are *parabolic*. One undesirable consequence of the parabolic nature of hydrodynamic equations is that perturbations propagate instantaneously.[9]

3. The presence of the second derivative in time implies that the initial condition $\rho(x, t = 0) = \rho_0(x)$ must be supplemented by an initial condition on the first derivative $(\partial \rho / \partial t)\big|_{t=0}$.

To solve the telegraph equation (3.30), let $\rho(k, s)$ be the combined Laplace-Fourier transform of $\rho(x, t)$, and let's consider the symmetric initial condition

$$f(x, v, t = 0) = \delta(x) \frac{\delta(v - v_0) + \delta(v + v_0)}{2} . \tag{3.32}$$

Thus $\rho_+(x, t = 0) = \frac{1}{2}\rho_-(x, t = 0) = \delta(x)$ and $J(x, t = 0) = 0$, so that the initial condition for the total density becomes $\rho(x, t = 0) = \delta(x)$, while Eq. (3.29b) gives $\rho_t(x, t = 0) = 0$, where the subscript denotes partial derivative. In the Fourier domain, these initial conditions become $\rho(k, t = 0) = 1$, $\rho_t(k, t = 0) = 0$. Using these results, the Laplace-Fourier

---

[9] This unphysical feature of hydrodynamic equations is usually ignored as a small price to pay for the vast success of the Navier–Stokes equation. Higher-order hydrodynamic equations (like the Burnett equations), that should be more accurate, are plagued by instabilities that originate from asymmetry between space and time coordinates – first-order derivative in time versus higher-order in spatial coordinates.

transform of the total density is

$$\rho(k, s) = \frac{1 + \frac{1}{2}\tau s}{Dk^2 + s + \frac{1}{2}\tau s^2}. \tag{3.33}$$

Inverting this transform then gives

$$\rho(x, t) = \frac{1}{2} e^{-t/\tau} \delta(v_0 t - |x|) + \frac{1}{2v_0\tau} e^{-t/\tau} \left[I_0(\xi) + \frac{t}{\tau\xi} I_1(\xi) \right] \theta(v_0 t - |x|), \tag{3.34}$$

where $\xi = \tau^{-1}\sqrt{t^2 - (x/v_0)^2}$ and I_n is the modified Bessel function of order n. The first term on the right-hand side of Eq. (3.34) accounts for particles that have not yet experienced any collision, while the second term accounts for all other particles; the latter must remain inside the interval $[-v_0 t, v_0 t]$, which is imposed by the Heaviside step function θ. In the long-time limit, the asymptotic expansion of the Bessel functions leads to (problem 3.6)

$$\rho(x, t) = \frac{1}{\sqrt{4\pi Dt}} e^{-x^2/4Dt} \, \theta(v_0 t - |x|), \tag{3.35}$$

with $D = v_0^2\tau/2$. Thus we recover classical diffusion in one dimension.

3.3 Lorentz gas in an external field

While Lorentz models were introduced to describe the electrical conductivity in metals, ironically, their application to the motion of electrons in a uniform electric field does *not* lead to Ohm's law, $\mathbf{J} = \sigma\mathbf{E}$, that relates the current density $\mathbf{J}$ to the electric field $\mathbf{E}$. The reason for this disconnect is simple – the Lorentz model contains no dissipation, so that charged particles preferentially move along the field and their energy increases without bound. The influence of a magnetic field can be even more profound because unperturbed cyclotron orbits can exist.

Lorentz gas in an electric field

Let's now understand the counterintuitive behaviors of the field-driven Lorentz gas. To compute the electrical conductivity, Lorentz used the linear Boltzmann equation

$$\left(\frac{\partial}{\partial t} + \mathbf{v} \cdot \frac{\partial}{\partial \mathbf{r}} - e\mathbf{E} \cdot \frac{\partial}{\partial \mathbf{v}} \right) f(\mathbf{r}, \mathbf{v}, t) = \frac{|\mathbf{v}|}{\ell} \left[\left(\int \frac{d\widehat{\mathbf{v}}}{4\pi} f(\mathbf{r}, \mathbf{v}, t) \right) - f(\mathbf{r}, \mathbf{v}, t) \right], \tag{3.36}$$

and proposed an approximate stationary solution of (3.36) by a perturbative expansion around the Maxwell–Boltzmann distribution; a form that should apply when the electric field is sufficiently small. However, the starting point of Lorentz'a analysis is erroneous.[10] If the

---

[10] Lorentz's paper led to the Lorentz models, which played an essential role in the development of kinetic theory and also triggered much mathematical research in the development of ergodic theory. Here is an example that a creative paper with an erroneous result can be more valuable than hundreds of correct papers.

scattering is elastic (no dissipation, as Lorentz assumed), then an electron will necessarily gain energy and a stationary state cannot exist. Scattering still plays a significant role, however; we shall see that the typical velocity grows as $t^{1/3}$ rather than the linear growth in time that characterizes a freely accelerated electron.

Before studying the consequences of the BL equation (3.36) in three dimensions, let's look at the one-dimensional version[11]

$$\left[\frac{\partial}{\partial t} + v\frac{\partial}{\partial x} + \mathcal{E}\frac{\partial}{\partial v}\right] f(x, v, t) = -\frac{|v|}{\ell} f^a(x, v, t). \tag{3.37}$$

Here $\mathcal{E}$ is the acceleration ($\mathcal{E} = -eE$ for an electron with charge $-e$ in an electric field E) and

$$f^s(x, v, t) = \frac{1}{2}[f(x, v, t) + f(x, -v, t)],$$

$$f^a(x, v, t) = \frac{1}{2}[f(x, v, t) - f(x, -v, t)] \tag{3.38}$$

are, respectively, the symmetric and antisymmetric components of the probability density.

Although the speed is no longer conserved, energy is conserved. For an electron that starts at $x = x_0$ with velocity $v = v_0$, energy conservation gives

$$x - x_0 = \frac{v^2 - v_0^2}{2\mathcal{E}}. \tag{3.39}$$

Hence the probability density reduces to the velocity distribution:

$$f(x, v, t | x_0, v_0) = F(v, t | v_0)\, \delta\left(x - x_0 - \frac{v^2 - v_0^2}{2\mathcal{E}}\right).$$

Using this factorization the velocity distribution satisfies

$$\left(\frac{\partial}{\partial t} + \mathcal{E}\frac{\partial}{\partial v}\right) F + \frac{|v|}{\ell} F^a = 0, \tag{3.40}$$

or, by separating the symmetric and antisymmetric pieces of this equation:

$$\left(\frac{\partial}{\partial t} + \frac{|v|}{\ell}\right) F^a + \mathcal{E}\frac{\partial}{\partial v} F^s = 0,$$

$$\frac{\partial}{\partial t} F^s + \mathcal{E}\frac{\partial}{\partial v} F^a = 0. \tag{3.41}$$

For the canonical initial condition that a particle starts at rest, $F(v, 0) = \delta(v)$ (equivalently, $F^s(v, 0) = \delta(v)$ and $F^a(v, 0) = 0$), the exact solution to Eqs (3.41) can be written in terms of the Airy function. However, this solution does not provide intuition for two key features of the field-driven Lorentz gas – the $t^{1/3}$ growth of the velocity, and the asymptotic velocity distribution. Instead of presenting this exact solution, we give a qualitative explanation for these two features.

---

[11] We write the collision frequency as $|v|/\ell$ rather than $1/\tau$, as the speed is no longer constant.

Random walk argument for velocity growth

For convenience, let us assume that the scatterers are on a lattice of spacing ℓ and that an electron passes through a scatterer with probability $1/2$ and reflects backward with probability $1/2$. Let x_n and v_n be the position and velocity of the electron after the nth scattering event. Using (3.39) we have

$$v_{n+1}^2 - v_n^2 = 2\mathcal{E}(x_{n+1} - x_n) = \pm 2\mathcal{E}\ell . \tag{3.42}$$

The quantity v_n^2 therefore undergoes a random walk, so that the average of its square grows linearly with the number of steps, $\langle v_n^4 \rangle \sim (\mathcal{E}\ell)^2 n$. Thus the mean velocity grows as $\overline{v} \sim (\mathcal{E}\ell)^{1/2} n^{1/4}$ with the number of steps. To express the growth of $\overline{v}$ with time, we express the time increment between successive collisions as

$$t_{n+1} - t_n \sim \frac{\ell}{v_n} \sim \left(\frac{\ell}{\mathcal{E}}\right)^{1/2} \frac{1}{n^{1/4}} .$$

Therefore the time until the nth collision is

$$t_n = \sum_{k=0}^{n-1}[t_{k+1} - t_k] \sim \sqrt{\frac{\ell}{\mathcal{E}}} \sum_{k<n} k^{-1/4} \sim \sqrt{\frac{\ell}{\mathcal{E}}}\, n^{3/4} ,$$

leading to $n \sim (\mathcal{E}t^2/\ell)^{2/3}$. Eliminating n in favor of t, we obtain

$$\overline{v} \sim (\mathcal{E}^2 \ell\, t)^{1/3} . \tag{3.43}$$

We emphasize that the mean speed (3.43) is distinct from the drift velocity. To estimate the latter, it is helpful to think of the field-driven Lorentz gas as an unbiased random walk (since scattering is isotropic) on the half-line $x > 0$, with a reflecting boundary at $x = 0$. The typical position of the electron divided by the time gives the drift velocity. For an electron that starts from rest at the origin, its position and speed are related by $x = v^2/2\mathcal{E}$. Since $\overline{v} \sim (\mathcal{E}^2 \ell\, t)^{1/3}$, the typical position of the electron is $x \sim (\mathcal{E}\ell^2 t^2)^{1/3}$, from which we infer a drift velocity $v_{\text{drift}} \sim (\mathcal{E}\ell^2/t)^{1/3}$.

Thus the drift velocity decreases with time even though the mean speed increases without bound. Since the current is proportional to the drift velocity, $J \sim E^{1/3} t^{-1/3}$. This result contradicts Ohm's law, according to which the current should be stationary and proportional to E.

Asymptotic velocity distribution

Scaling, together with symmetry arguments, provide an appealing way to deduce the asymptotic velocity distribution. In the long-time limit, we expect that the symmetric part of the velocity distribution approaches the scaling form

$$F^s(v,t) = t^{-1/3}\Phi(u), \qquad u = vt^{-1/3} . \tag{3.44}$$

This form is consistent with the growth law $v \sim t^{1/3}$ in Eq. (3.43), and the prefactor ensures the normalization $\int dv \, F^s(v,t) = 1$. (In terms of the scaled distribution, the normalization becomes $\int du \, \Phi(u) = 1$.) The isotropy of the symmetric part of the velocity distribution arises because the average drift velocity decreases while the mean velocity increases.

From the second of Eqs (3.41) we postulate the scaling form of the asymmetric part,

$$F^a(v,t) = t^{-1} u \Psi(u), \tag{3.45}$$

with $\Psi(u)$ an even function so that the scaling form is manifestly antisymmetric. Substituting (3.44)–(3.45) into Eqs (3.41) we obtain

$$\frac{|u| u}{\ell} \Psi + \mathcal{E} \Phi' = 0, \tag{3.46a}$$

$$-\frac{1}{3} [\Phi + u\Phi'] + \mathcal{E} [\Psi + u\Psi'] = 0, \tag{3.46b}$$

where the prime denotes differentiation with respect to u. Integrating (3.46b) gives $\Psi = \Phi/(3\mathcal{E})$. Therefore (3.46a) becomes $\Phi'/\Phi = -u|u|/(3\ell\mathcal{E}^2)$, from which we can calculate $\Phi(u)$. Assembling everything we obtain

$$F^s(v,t) = \frac{1}{2\Gamma(1/3)} \left(\frac{3}{\ell\mathcal{E}^2 t} \right)^{1/3} e^{-|v|^3/9\ell\mathcal{E}^2 t}, \tag{3.47a}$$

$$F^a(v,t) = \frac{v}{3\mathcal{E}t} F^s(v,t). \tag{3.47b}$$

These results agree with the limiting asymptotic velocity distribution that arises from the full solution of Eqs (3.41) with the initial condition $F(v,0) = \delta(v)$.

Lorentz gas in a magnetic field

We know that, in a constant uniform magnetic field $\mathbf{B}$, an electron moves along a spiral, with a circular projection on the plane perpendicular to the field. The velocity component $\mathbf{v}_\parallel$ of an electron parallel to the field remains constant, while the velocity component $\mathbf{v}_\perp$ rotates with the cyclotron frequency $\omega = eB/mc$. (Here m is the electron mass and c is the speed of light.) The electrons (with charge $-e$) experience a Lorentz force so that their acceleration is $\mathbf{F}/m = \omega \widehat{\mathbf{B}} \times \mathbf{v}$, where $\widehat{\mathbf{B}}$ is a unit vector in the direction of the magnetic field: $\mathbf{B} = B\widehat{\mathbf{B}}$. The probability density $f(\mathbf{r}, \mathbf{v}, t)$ now satisfies the BL equation

$$\left(\frac{\partial}{\partial t} + \mathbf{v} \cdot \frac{\partial}{\partial \mathbf{r}} + \omega \left(\widehat{\mathbf{B}} \times \mathbf{v} \right) \cdot \frac{\partial}{\partial \mathbf{v}} \right) f(\mathbf{r}, \mathbf{v}, t) = \frac{1}{\tau} \left[\left(\int \frac{d\widehat{\mathbf{v}}}{4\pi} f(\mathbf{r}, \mathbf{v}, t) \right) - f(\mathbf{r}, \mathbf{v}, t) \right]. \tag{3.48}$$

The speed of the electron does not change during the cyclotron motion and it also remains the same after an elastic collision with a scatterer (Fig. 3.2); therefore the collision frequency $1/\tau = v/\lambda$ remains constant. In this respect, the BL equation in the magnetic field is similar to the BL without an external force. This fact suggests analyzing Eq. (3.48) using the same approach as in Eq. (3.7).
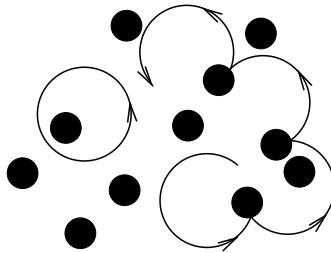
Fig. 3.2. Lorentz gas of circular scatterers in a magnetic field. Shown is a closed cyclotron orbit in which there is no scattering, and a portion of a trajectory that undergoes multiple scattering.

Integrating (3.48) over the spatial coordinates, we find that the velocity distribution $F(\mathbf{v}, t)$ satisfies (see Eq. (3.10))

$$\left(\frac{\partial}{\partial t} + \omega \left(\widehat{\mathbf{B}} \times \mathbf{v}\right) \cdot \frac{\partial}{\partial \mathbf{v}}\right) F(\mathbf{v}, t) = \frac{1}{\tau} \left([\mathbb{P}F](v, t) - F(\mathbf{v}, t)\right). \tag{3.49}$$

Applying the projector operator $\mathbb{P}$ to (3.49) we arrive at (3.12), the same equation as in the force-free case. Thus the spherical part of the velocity distribution does not change. The anisotropic part $F^a(\mathbf{v}, t) = F(\mathbf{v}, t) - [\mathbb{P}F](v, t)$ obeys

$$\left(\frac{\partial}{\partial t} + \omega \left(\widehat{\mathbf{B}} \times \mathbf{v}\right) \cdot \frac{\partial}{\partial \mathbf{v}}\right) F^a(\mathbf{v}, t) = -\tau^{-1} F^a(\mathbf{v}, t) \tag{3.50}$$

and it decays to zero exponentially quickly (with relaxation time of the order of τ). This assertion can be understood without calculation; it suffices to rewrite (3.50) as

$$\left(\frac{\partial}{\partial t} + \omega \left(\widehat{\mathbf{B}} \times \mathbf{v}\right) \cdot \frac{\partial}{\partial \mathbf{v}}\right) \Phi(\mathbf{v}, t) = 0, \qquad \Phi(\mathbf{v}, t) = e^{t/\tau} F^a(\mathbf{v}, t) \tag{3.51}$$

to see the emergence of an exponential decay with the relaxation time $\tau = \lambda/v$.

The analysis of the influence of the magnetic field on the relaxation of an initial density inhomogeneity is more challenging than in the no-field case. Indeed, the latter situation is intrinsically isotropic – even if the initial conditions are anisotropic, the long-time state is isotropic. One consequence of this isotropy is that diffusive motion is characterized by a *scalar* diffusion coefficient. In the presence of a magnetic field, however, the resulting anisotropic long-time diffusive motion is governed by a diffusion *tensor*. If the magnetic field points in the z direction, then symmetry between x and y and the lack of coupling between the motion in the x–y plane and along the z axis tells us that the diffusion tensor has the form (problem 3.8)

$$\begin{pmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{yx} & D_{yy} & D_{yz} \\ D_{zx} & D_{zy} & D_{zz} \end{pmatrix} = \begin{pmatrix} D_\perp & D_{xy} & 0 \\ D_{yx} & D_\perp & 0 \\ 0 & 0 & D_\parallel \end{pmatrix}. \tag{3.52}$$

A physically obvious, yet startling, feature of the Lorentz model in a uniform magnetic field is that the BL equation is sometimes inapplicable. In three dimensions, the electron has,

with probability one, a non-zero velocity component parallel to the magnetic field. Hence an electron moving along a spiral will surely collide with a scatterer if they are randomly distributed; the BL equation therefore provides a proper description in three dimensions.[12] For electrons confined to a plane with the magnetic field perpendicular to this plane, the situation is drastically different. Now closed cyclotron orbits can exist that do not intersect with any scatterer (Fig. 3.2).[13] The existence of freely circulating electrons implies that the standard Boltzmann description, in which scattering occurs at a constant rate, is invalid.

The two-dimensional Lorentz gas in a uniform transverse magnetic field may seem exotic; however, two-dimensional electron films are routinely manufactured with high perfection in GaAs heterostructures. Studies of such systems in strong magnetic fields is an active research direction – apart from being in practically every semiconductor device, two-dimensional electron systems gave rise to interesting physics, such as the quantum Hall effect. The quantum nature of electrons is crucial in most of these systems; nevertheless, even on the classical level the two-dimensional Lorentz gas in a transverse magnetic field is interesting and challenging.

3.4 Collisional impact

The challenge of understanding hard-sphere gases motivated the much simpler Lorentz model of the previous two sections. In a similar spirit, let us consider the response of a static hard-sphere gas to a single moving particle. As we shall see, this setting has a close connection with the propagation of a blast wave following an explosion. After the first collision, two particles are moving and these, in turn, collide with other particles, ultimately leading to a remarkably symmetrical cascade of moving particles (Fig. 3.3). We wish to understand the basic properties of this cascade.

The sudden injection of energy into the zero-temperature gas can be regarded as an explosion, and the collision cascade is analogous to the shock wave that propagates outward from the initial detonation. The basic properties of a shock wave, such as its propagation velocity, and the radial dependences of the density and temperature profiles behind the wave, are well understood from hydrodynamic theory. As we now show, the hard-sphere collision cascade shares many of the quantitative properties of this shock wave.

Two natural characteristics of the cascade are the number $N(t)$ of moving particles at time t, and the total number of collisions $C(t)$ up to time t. As in the blast wave, we may use dimensional analysis to determine these two quantities with minimal labor. The number of moving particles can depend on time t, the density n, the spatial dimension d, the initial velocity v_0 of the incident particle, and the particle radii a: $N(t) = N[t, v_0, a, n, d]$. Dimensional analysis tells us that $N(t)$ can depend on the dimensionless time $T = n^{1/d} v_0 t$ and two other dimensionless parameters, e.g. $\rho = nV_d a^d$, which is the volume fraction

---

[12] We assume that the standard requirements – dilute limit, randomly located scatterers, etc. – are satisfied.

[13] More precisely, a free cyclotron orbit occurs with probability $e^{-2\pi R/\lambda} = e^{-2\pi/(\omega\tau)}$; $R = v/\omega$ is the radius of this orbit.
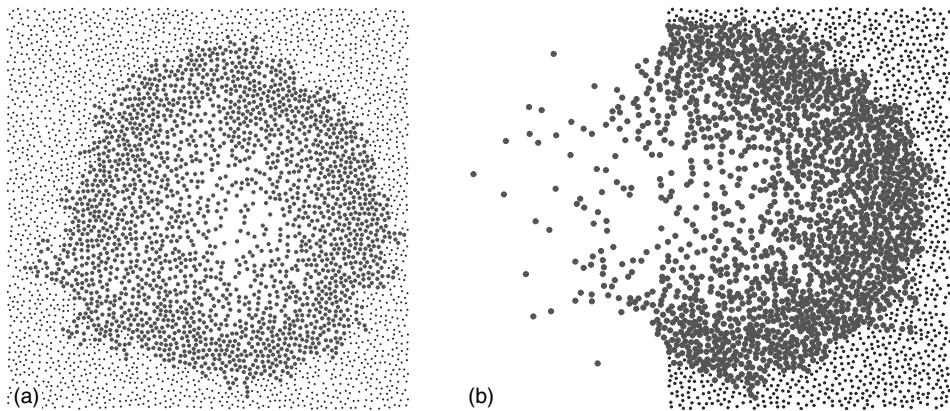
Fig. 3.3. (a) Collision cascade in two dimensions when a single particle suddenly starts moving inside a zero-temperature gas. Moving spheres are shown actual size and stationary ones at half radius. (b) The same when the gas occupies the half-space $x \geq 0$ and a moving particle is incident from the left. Courtesy of T. Antal.

occupied by particles[14] and d. Therefore

$$N = N(T, \rho, d). \tag{3.53}$$

A self-similar nature of the collision cascade suggests that $N(t)$ should have the power-law time dependence

$$N \sim T^{\xi}, \tag{3.54}$$

with exponent ξ that potentially depends on ρ and d.

Let us understand how the number of moving particles grows with time by elementary kinetic theory. From Fig. 3.3, only particles within a shell whose thickness is of the order of one mean free path $\ell \sim v\tau \sim 1/(na^{d-1})$ from the edge of the cascade can recruit additional moving particles in a collision time τ. If the energy in the cascade is shared roughly equally among the moving particles, each particle in the cascade has energy that is of the order of N^{-1} times the initial energy, and hence $v \sim N^{-1/2}$. Since the collision cascade (Fig. 3.3) is approximately a ball of (growing) radius R, the number of moving particles increases with time according to

$$\frac{dN}{dt} \sim n\, v R^{d-1} \sim nv \left(\frac{N}{n}\right)^{(d-1)/d}. \tag{3.55}$$

Using $v \sim N^{-1/2}$ and solving the resulting differential equation confirms the power-law growth (3.54) and yields $\xi = 2d/(d+2)$. Surprisingly, the exponent ξ does not depend on the volume fraction ρ. A similar line of reasoning can be extended to determine the total number of collisions that have occurred up to time t (problem 3.9).

[14] $V_d = \pi^{d/2}/\Gamma(1 + d/2)$ is the volume of the unit ball in d dimensions.

The properties of this collision cascade can also be determined by a hydrodynamically based dimensional analysis.[15] The perturbation caused by a single moving particle can be viewed as the instantaneous release of energy $E = \frac{1}{2}mv_0^2$. This point energy source is akin to detonating an explosion, and the latter problem has an elegant scaling solution, as long as the pressure behind the shock wave greatly exceeds that in front. This condition is always satisfied in our case, as the exterior gas has zero pressure. In this infinite Mach number limit, the radius $R(t)$ of the shock wave can only depend on the energy release E, the mass density nm, where m is the particle mass, and the time t. Therefore the radius must scale as

$$R(t) \sim (Et^2/nm)^{1/(d+2)} \; , \tag{3.56}$$

as this is the only variable combination with units of length.[16] This result for $R(t)$, in conjunction with $E = \frac{1}{2}mv_0^2$ and $N = nV_dR^d$, reproduces (3.54) that was obtained by kinetic theory.

Consider finally the situation where a particle is normally incident on a zero-temperature gas in the half-space $x \geq 0$ (Fig. 3.3). Again a collision cascade propagates in the gas in a manner similar to that of the bulk system, but some particles are ultimately ejected backward. Even though visually this backsplatter consists of a small number of particles, eventually all the energy is transferred to this backsplatter. Paradoxically, this transfer of energy to the backsplatter occurs very slowly. A heuristic argument suggests that the energy in the initially occupied half-space decays to zero as a power law in time, with an as yet unknown exponent (problem 3.10).

3.5 Maxwell molecules and very hard particles

The BL equation faithfully represents many properties of the BE, but there is one essential missing element – the BE is *nonlinear* whereas the BL equation is linear. As alluded to earlier in this chapter, there has been a long history of making simplifying modifications of the full nonlinear BE to obtain a tractable theory. Here we discuss two such examples that have had considerable utility.

- *Maxwell molecules* interact via the potential $U(r) = A/r^4$. The BE for this model greatly simplifies because the relative velocity inside the collision integral of Eq. (3.6) is replaced by a constant. To understand why the collision rate becomes independent of the relative velocity u, suppose that the intermolecular potential is $U(r) = A/r^\nu$. We then estimate the closest approach of two molecules by equating the kinetic and potential energies, $Ar^{-\nu} \sim mu^2$, leading to $r \sim u^{-2/\nu}$. Then the cross-section $\sigma \sim r^{d-1}$ has the velocity

---

[15] This was the legendary argument given by G. I. Taylor, who estimated the yield of the first A-bomb explosion from time-resolved photographs of the radius of the blast wave as a function of time.

[16] What is not determined by this dimensional analysis is the prefactor in Eq. (3.56); this requires solving the full hydrodynamic equations of motion.

dependence $u^{-2(d-1)/v}$, and the overall collision rate scales as

$$u\sigma \sim u\,r^{d-1} \sim u^{1-2(d-1)/v}\,.$$

This rate is independent of the relative velocity when $v = 2(d-1)$. In three dimensions, the r^{-4} Maxwell intermolecular potential arises. While this kind of interaction has not been found in Nature (for example, the van der Waals potential between neutral molecules scales as r^{-6}), a virtue of the Maxwell model is that it is analytically tractable.

- *Very hard particles* have an overall collision rate that is proportional to the *square* of the relative velocity between two molecules. Because of this quadratic velocity dependence, the interaction is "harder" than the hard-sphere interaction of classical kinetic theory, while the Maxwell model is "softer" because of the velocity-independent interaction. Maxwell molecules and very hard particles can therefore be viewed as bounds on classical kinetic theory.

The Boltzmann equations for Maxwell molecules and for very hard particles are still quite formidable and it is not obvious *a priori* why they should be simpler than the BE for hard spheres. We can give some hints as to why these models are easier to deal with by estimating the temperature dependence of the transport coefficients. For the diffusion constant, we start with the kinetic theory relation $D \sim v\ell \sim v/(n\sigma)$, with thermal velocity $v \sim \sqrt{T}$. For Maxwell molecules, the collision rate $nv\sigma$ is postulated to be independent of the relative velocity; consequently, σ must be proportional to v^{-1}. Conversely, for very hard particles, the collision rate is postulated to be quadratic in velocity, so that $\sigma \sim v$. Thus we infer the temperature dependences:

$$D \sim \frac{v}{\sigma} \sim \begin{cases} v^2 \sim T, & \text{Maxwell molecules}\,, \\ v \sim \sqrt{T}, & \text{hard spheres}\,, \\ 1, & \text{very hard particles}\,. \end{cases} \tag{3.57}$$

Other transport coefficients (such as the kinematic viscosity) exhibit the same temperature dependence. The non-analytic dependence of D on temperature for hard spheres suggests that computing transport coefficients for this basic model could be challenging. Indeed, the exact transport coefficients are unknown for a hard-sphere gas despite the efforts of Boltzmann, Hilbert, Chapman, Enskog, and others. In contrast, Maxwell himself calculated the transport coefficients for Maxwell molecules.[17]

Application to ballistic annihilation

The utility of Maxwell molecules and very hard particles is particularly striking in irreversible ballistic annihilation, where particles move at constant velocity and annihilate upon colliding. Despite the simplicity of this process, it is not solvable for a hard-sphere

---

[17] Again, Maxwell introduced his model in 1867, before the Boltzmann equation was first written, and much before the Chapman–Enskog scheme had been devised.

gas, but it is solvable for Maxwell molecules and very hard particles. Within the Boltz-mann framework, the velocity distribution of a spatially homogeneous gas, $P(\mathbf{v}, t)$, evolves according to

$$\frac{\partial P(\mathbf{v}, t)}{\partial t} = -P(\mathbf{v}, t) \int g \, \sigma \, P(\mathbf{w}, t) \, d\mathbf{w} . \tag{3.58}$$

In the spirit of the Boltzmann approach, the probability that two particles with velocities $\mathbf{v}$ and $\mathbf{w}$ collide factorizes into a product of single-particle densities. The remaining kinematic factors consist of the relative velocity $g = |\mathbf{v} - \mathbf{w}|$ times the cross-section σ that we can absorb into a rescaled time variable. The right-hand side consists of only a loss term, since there is only annihilation. In one dimension Eq. (3.58) becomes

$$\frac{\partial P(v, t)}{\partial t} = -P(v, t) \int_{-\infty}^{\infty} dw \, |v - w| \, P(w, t) . \tag{3.59}$$

This innocent-looking equation has not been solved, and the only results have come from numerical simulations that indicate that the density and the average speed decay algebraically with time,

$$n(t) = \int_{-\infty}^{\infty} dw \, P(w, t) \sim t^{-\alpha}, \qquad v_{\rm rms}(t) = \left[\frac{1}{n(t)} \int_{-\infty}^{\infty} dw \, w^2 \, P(w, t) \right]^{1/2} \sim t^{-\beta}. \tag{3.60}$$

The exponents α and β depend on the small-v tail of the initial velocity distribution $P_0(v)$. (Only one of the two exponents is independent, since $\alpha + \beta = 1$; see problem 3.15.) For example, for isotropic distributions that are flat near the origin, the decay exponent is $\beta_{\rm flat} \doteq 0.230472$; for the initial distribution linearly vanishing at the origin, $P_0(v) \sim |v|$, the decay exponent is $\beta_{\rm linear} \doteq 0.166649$.[18] On the other hand, for Maxwell molecules and for very hard particles ballistic annihilation can be solved completely.

Example 3.1. *Maxwell molecules.* Since molecules interact at a rate that is independent of the velocity, the mean molecular velocity does not change. We therefore replace the factor $g = |\mathbf{v} - \mathbf{w}|$ in (3.58) by a constant u, with units of velocity for dimensional correctness, to recast Eq. (3.58) as

$$\frac{\partial P(\mathbf{v}, t)}{\partial t} = -P(\mathbf{v}, t) \, n(t) \, u\sigma . \tag{3.61}$$

Integrating over velocities gives $\dot{n} = -n^2 u\sigma$, which is trivially soluble, and using this result in (3.61) gives

$$n(t) = \frac{n(0)}{1 + n(0) \, u\sigma \, t}, \qquad\qquad P(\mathbf{v}, t) = \frac{P_0(\mathbf{v})}{1 + n(0) \, u\sigma \, t} . \tag{3.62}$$

---

[18] A direct simulation of ballistic annihilation only gives exponents with less than two-digit accuracy; the more accurate values quoted above are obtained by substituting the scaling form $P(v, t) \sim t^{\beta - \alpha} f(vt^\beta)$ into Eq. (3.59) and then numerically solving a resulting nonlinear eigenvalue equation.

The resulting decay exponents are $\alpha = 1$ and $\beta = 0$; these are universal and trivial in the Maxwell model.

Example 3.2. *Very hard particles.* Here, dimensionally faithful modeling means replacing g in Eq. (3.58) by $g^2/v_{\text{rms}}(t)$. Absorbing the factor $1/v_{\text{rms}}(t)$ into the time variable,

$$\tau = \int_0^t \frac{\sigma}{v_{\text{rms}}(t')}\, dt' \,, \tag{3.63}$$

we recast the BE as

$$\frac{\partial P(\mathbf{v}, \tau)}{\partial \tau} = -P(\mathbf{v}, \tau) \int |(\mathbf{v} - \mathbf{w})^2| P(\mathbf{w}, \tau)\, d\mathbf{w} = -P(\mathbf{v}, \tau)\, n(\tau)[v^2 + v_{\text{rms}}^2(\tau)], \tag{3.64}$$

where we use $|(\mathbf{v} - \mathbf{w})^2| = v^2 + w^2 - 2\mathbf{v} \cdot \mathbf{w}$ and $\int d\mathbf{w}\, \mathbf{w} P(\mathbf{w}, \tau) \equiv 0$ (zero-momentum reference frame) to write the second line. The solution to Eq. (3.64) is

$$P(\mathbf{v}, \tau) = P_0(\mathbf{v})\, e^{-v^2 A(\tau) - B(\tau)} \,, \tag{3.65}$$

with

$$A(\tau) = \int_0^\tau d\tau'\, n(\tau'), \qquad B(\tau) = \int_0^\tau d\tau'\, n(\tau') v_{\text{rms}}^2(\tau') \,. \tag{3.66}$$

To obtain explicit results, consider the generic one-parameter family of initial velocity distributions:

$$P_0(\mathbf{v}) = \frac{2}{\Omega_d\, \Gamma[(\mu + d)/2]}\, v^\mu e^{-v^2}, \tag{3.67}$$

where $\Omega_d = 2\pi^{d/2}/\Gamma(d/2)$ is the surface area of the unit sphere in d dimensions, and the prefactor of $P_0(\mathbf{v})$ is chosen to set the initial density to one. Integrating (3.65) we obtain

$$n(\tau) = \int d\mathbf{v}\, P(\mathbf{v}, \tau) = \frac{2}{\Omega_d\, \Gamma[(\mu + d)/2]} \int d\mathbf{v}\, v^\mu\, e^{-v^2[1 + A(\tau)] - B(\tau)}$$

$$= [1 + A(\tau)]^{-(\mu + d)/2}\, e^{-B(\tau)} \,.$$

Using this together with Eq. (3.66) we get

$$\frac{dA}{d\tau} = n(\tau) = [1 + A]^{-(\mu + d)/2}\, e^{-B} \,.$$

Similarly, multiplying (3.65) by v^2 and integrating, and comparing with the derivative of the second equation in (3.66), we obtain

$$\frac{dB}{d\tau} = \frac{\mu + d}{2}\, [1 + A]^{-(\mu + d + 2)/2}\, e^{-B} \,.$$

Solving these equations for A and B, subject to $A(0) = B(0) = 0$, we find A and B, from which we compute the density and the mean speed as

$$n = [1 + (\mu + d + 1)\tau]^{-(\mu+d)/(\mu+d+1)},$$

$$v_{\text{rms}} = \sqrt{\frac{\mu + d}{2}} \, [1 + (\mu + d + 1)\tau]^{-1/[2(\mu+d+1)]} \, . \tag{3.68}$$

Using Eq. (3.63) to eliminate τ in favor of the time, we re-express the density and the mean speed in terms of the time and thereby obtain the density and velocity decay exponents:
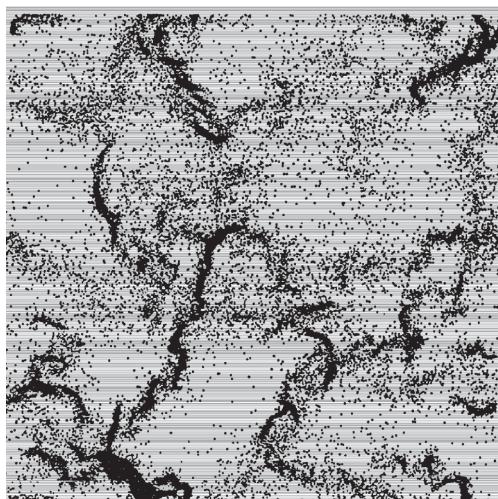
$$\alpha = \frac{2(\mu + d)}{1 + 2(\mu + d)}, \qquad \beta = \frac{1}{1 + 2(\mu + d)} \, . \tag{3.69}$$

3.6 Inelastic gases

Kinetic theory traditionally deals with gases whose molecules interact via elastic collisions. In contrast, collisions between macroscopic particles are (almost always) inelastic. Many new and unexpected phenomena arise – perhaps the most spectacular are the large-scale clustering and the *inelastic collapse* of a freely evolving inelastic gas (Fig. 3.4).

The basic parameter that characterizes inelastic collisions is the *restitution coefficient r*. In one dimension, this quantity is defined as

$$r = -\frac{v_{12}}{u_{12}} \, , \tag{3.70}$$

Fig. 3.4. Inelastic clustering for a gas of 40000 particles with restitution coefficient 0.6. The area fraction is 0.05 and there are on average 500 collisions per particle. (From I. Goldhirsch and G. Zanetti, *Phys. Rev. Lett.* 73, 1619 (1993), with permission.)

where $u_{12} = u_1 - u_2$ and $v_{12} = v_1 - v_2$ are the relative velocities of two particles before and after a collision. The restitution coefficient lies within the bounds $0 \leq r \leq 1$, with the lower bound corresponding to totally inelastic collisions and the upper bound to elastic collisions. If the masses of the colliding particles are equal, momentum conservation, $v_1 + v_2 = u_1 + u_2$, together with the definition (3.70) determine the post-collision velocities (v_1, v_2) in terms of the pre-collision velocities (u_1, u_2):

$$
\begin{aligned}
v_1 &= \frac{1}{2}(1 - r)u_1 + \frac{1}{2}(1 + r)u_2 \equiv \epsilon u_1 + (1 - \epsilon)u_2, \\
v_2 &= \frac{1}{2}(1 + r)u_1 + \frac{1}{2}(1 + r)u_2 \equiv (1 - \epsilon)u_1 + \epsilon u_2.
\end{aligned}
\tag{3.71}
$$

For later convenience we parameterize the kinematics by the dissipation parameter $\epsilon \equiv \frac{1}{2}(1 - r)$. Setting the particle mass to one, the energy loss in a collision is

$$
\Delta E = \frac{1}{2}(u_1^2 + u_2^2) - \frac{1}{2}(v_1^2 + v_2^2) = -\epsilon(1 - \epsilon)(u_1 - u_2)^2.
\tag{3.72}
$$

The energy loss is maximal for completely inelastic collisions ($r = 0, \epsilon = 1/2$) and vanishes for elastic collisions ($r = 1, \epsilon = 0$).[19]

The mean kinetic energy of particles that undergo inelastic collisions is called the *granular temperature*, or, customarily, the temperature. Because of inelastic collisions, the temperature decreases with time. What is the time dependence of this cooling? Using elementary kinetic theory, we can determine this time dependence under the assumption that the gas remains spatially homogeneous. Homogeneity is a reasonable approximation at early times, but this naive assumption is erroneous in the long-time limit because of large-scale clustering.

Let's first analyze the outcome of a typical collision. From Eq. (3.72), the kinetic energy lost in an inelastic collision is $\Delta T = -\epsilon(1 - \epsilon)(\Delta v)^2 \sim -\epsilon(\Delta v)^2$, with Δv the relative velocity between colliding particles. We focus on the quasi-elastic limit $\epsilon \to 0$, where spatial homogeneity is a good approximation over a substantial time range. The typical time Δt between collisions is roughly $\ell/\Delta v$, with ℓ the mean free path. We assume that a single scale characterizes all velocities so that $\Delta v \sim v \sim \sqrt{T}$. Assembling these elements, we estimate that the temperature cools in time according to

$$
\frac{dT}{dt} \sim \frac{\Delta T}{\Delta t} \sim -\frac{\epsilon(\Delta v)^2}{\ell/\Delta v} \sim -\frac{\epsilon T^{3/2}}{\ell}.
$$

---

[19] In three dimensions, the relative velocity $\mathbf{u}_{12} = \mathbf{u}_1 - \mathbf{u}_2$ has normal and tangential parts,

$$
\mathbf{u}_{12} = \mathbf{g}_n + \mathbf{g}_t, \qquad \mathbf{g}_n = (\mathbf{u}_{12} \cdot \mathbf{n})\,\mathbf{n} \quad \text{and} \quad \mathbf{g}_t = \mathbf{u}_{12} - \mathbf{g}_n,
$$

where $\mathbf{n}$ is the unit vector that joins the centers of two particles at their point of collision. Generally, there are separate restitution coefficients for the normal and tangential velocities. This feature requires rotation to be accounted for, even if particles were initially not rotating. In the following, such rotational motion is ignored.

From this rate, the temperature decay is

$$T(t) = T_0 \left(1 + \frac{t}{t_0}\right)^{-2}, \qquad t_0 = \frac{\ell}{\epsilon\sqrt{T_0}}. \tag{3.73}$$

The gas remains effectively elastic, $T(t) \simeq T_0$, for $t \ll t_0$; this time range can be substantial if the collisional dissipation is weak. Beyond this time range, the temperature decreases algebraically in time, $T(t) \sim (\epsilon t/\ell)^{-2}$. This decay is *Haff's law*.

The arguments underlying Haff's law apply if the inelasticity is sufficiently small or if the system contains sufficiently few particles. Otherwise, large-scale density inhomogeneities develop, leading to deviations from Haff's law and to ultimate inelastic collapse.

Inelastic collapse in one dimension

To help understand inelastic collapse, consider the trivial example of a single inelastic particle that bounces on a horizontal plate. If the particle starts at the plate with an initial upward velocity v, the particle will hit the plate again after a time interval $2v/g$ (g is the gravitational acceleration). After the collision, the particle will move upward with velocity rv and the cycle repeats. The time required for an infinite number of collisions is

$$t_c = \frac{2v}{g}(1 + r + r^2 + r^3 + \cdots) = \frac{2v}{g}\frac{1}{1 - r},$$

so that the particle undergoes inelastic collapse, i.e. comes to rest after an infinite number of collisions that occur in a *finite* time.

Two free particles undergo at most one collision, so collapse is impossible in this case. However, for three identical particles on the line, inelastic collapse occurs if the restitution coefficient r satisfies the bound

$$r < r_c = 7 - 4\sqrt{3} \doteq 0.071796769. \tag{3.74}$$

To establish this bound, we label the particles as $1, 2, 3$ from left to right and assume (without loss of generality) that the first collision is 12. The subsequent collisions then alternate between 23 and 12. After the collision pair 12 and 23, the post-collision velocities v_i are related to the pre-collision velocities u_i by (from Eq. (3.71))

$$\begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = M_{23}M_{12}\begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} \qquad \text{with} \quad M_{12} = \begin{pmatrix} \epsilon & 1-\epsilon & 0 \\ 1-\epsilon & \epsilon & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$M_{23} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \epsilon & 1-\epsilon \\ 0 & 1-\epsilon & \epsilon \end{pmatrix}. \tag{3.75}$$

After $2n$ collisions, the particle velocities are given by $\mathbf{v} = M^n \mathbf{u}_0$, where $M = M_{23}M_{12}$ and $\mathbf{u}_0$ is the initial velocity vector, while after $2n + 1$ collisions, the velocities are given by $\mathbf{v} = M_{12}M^n \mathbf{u}_0$. The evolution of the collision sequence is determined by the eigenvalues of the matrix M:

$$\lambda_{1,2} = \frac{1}{2}\left[(\epsilon^2 + 2\epsilon - 1) \pm \sqrt{(1 - 2\epsilon - \epsilon^2)^2 - 4(2\epsilon - 1)^2}\right], \quad \lambda_3 = 1.$$

When $\epsilon > \epsilon_c = 2\sqrt{3} - 3$, corresponding to restitution coefficient $r < r_c = 7 - 4\sqrt{3}$, then $\lambda_{1,2}$ are both real with absolute values less than 1. Consequently the particle velocities asymptotically decay as $[\max(|\lambda_1|, |\lambda_2|)]^n$, leading to ultimate inelastic collapse. Conversely for $r > r_c$, one of the eigenvalues is greater than 1 so that the collision sequence eventually terminates, after which the three particles mutually diverge from each other.

When $N > 3$, the collision sequence is not necessarily periodic and the matrix formulation given above no longer provides the exact value of the critical restitution coefficient r_c. We can instead give a heuristic argument for the dependence of r_c on N for large N. Consider a particle with speed 1 that is incident on an array of static, equally spaced particles when the restitution coefficient is nearly 1. After the first collision, the incident particle comes nearly to rest, while the target particle moves slightly slower than the initial particle. It is helpful to exchange the identities of these two particles, so that the world-line of particle 1 is merely deflected slightly as it "passes through" particle 2. This same pattern continues in subsequent collisions so that the initial particle world-line gradually bends as it penetrates the array (Fig. 3.5). Let $v(N)$ be the velocity of the initial trajectory after N collisions. From the collision rules (3.71) and accounting for the change in particle labeling, $v(1) = 1 - \epsilon$. Similarly, to first order in ϵ, $v(N) = (1 - \epsilon)^N \simeq 1 - N\epsilon$. For sufficiently large N, the initial particle momentum is eventually exhausted and inelastic collapse has occurred. Setting $v(N) = 0$ yields the critical dissipation parameter $\epsilon_c(N) \sim N^{-1}$, or critical restitution coefficient

$$1 - r_c(N) \sim \frac{1}{N}. \tag{3.76}$$

Thus inelastic collapse is inevitable for nearly elastic particles in the thermodynamic limit.
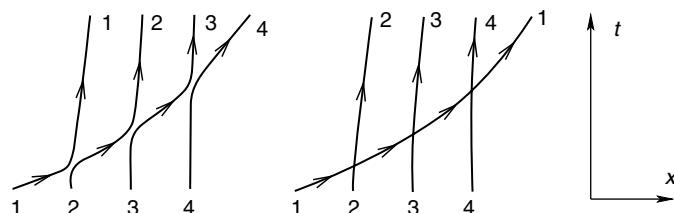
Space-time representation of a test particle as it collides inelastically and penetrates a static particle array. An equivalent representation is shown to the right in which particle labels are exchanged in each collision so that particle 1 penetrates the array but is decelerated by collisions.

Aside from this simple result, the asymptotic behavior of an inelastic gas in greater than one dimension is still incompletely understood. The existence of the mysteriously beautiful large-scale clustering shown in Fig. 3.4 invalidates any Boltzmann-like description of the gas. There is evidence that the asymptotic behavior is similar to that of ballistic agglomeration (see Section 3.7), where a gas undergoes totally inelastic collisions, a correspondence that is easiest to justify in one dimension. As a result of this equivalence, Haff's $T \sim t^{-2}$ cooling law applies only in an intermediate time regime, while the true asymptotic decay is given by $T \sim t^{-2/3}$ (see Eq. (3.94)), as predicted by the equivalence to agglomeration.

Inelastic Maxwell model in one dimension

Because it is not feasible to solve the inelastic collision dynamics of a real gas analytically, we turn to the Maxwell model in which the collision rate is independent of particle positions *and* velocities. Operationally in a simulation, we merely pick two particles at random, determine their post-collision velocities according to Eqs (3.71) in one dimension (or their generalization for $d > 1$), and update the time by a random variable whose mean is the inverse of the total number of particles. Because of the simplicity of this dynamics, we can explicitly derive an exact solution of the BE in one dimension, both for free cooling and for a driven gas.

Free cooling

The Boltzmann equation for the inelastic and spatially homogeneous Maxwell model is

$$\frac{\partial P(v,t)}{\partial t} = \sqrt{T} \iint P(u_1,t)P(u_2,t) \left\{ \delta \left[v - \epsilon u_1 - (1 - \epsilon)u_2 \right] - \delta(v - u_2) \right\} du_1 \, du_2 .$$

$$(3.77)$$

The delta functions encode the two ways in which the density of particles with speed v can change: creation occurs when particles with speeds ϵu_1 and $(1 - \epsilon)u_2$ collide (see Eq. (3.71)), or loss occurs when a particle with speed v interacts with any other particle. The leading factor of $\sqrt{T}$ arises from the rate at which two particles meet. This rate is independent of their relative speeds in the Maxwell model, but must have units of velocity to give a dimensionally correct equation. The only such scale in the system is the thermal velocity $\sqrt{T}$. Since the density is fixed, we use units in which the density is set to one, $\int P(v,t) \, dv = 1$, and we also set the total momentum to zero, $\int v P(v,t) \, dv = 0$.

Let us first study the low-order moments of the velocity distribution. We absorb the leading factor $\sqrt{T}$ into a rescaled time variable $\tau = \int_0^t \sqrt{T(t')} \, dt'$, multiply (3.77) by v^n, and integrate over v to obtain the evolution equations for the moments $M_n(\tau) = \int v^n P(v,\tau) \, dv$:

$$\frac{dM_n}{d\tau} + a_n M_n = \sum_{m=2}^{n-2} \binom{n}{m} \epsilon^m (1 - \epsilon)^{n-m} M_m M_{n-m} , \qquad (3.78)$$

with $a_n(\epsilon) = 1 - \epsilon^n - (1 - \epsilon)^n$. These equations may be solved recursively, starting with $M_0 = 1$ and $M_1 = 0$. The first few non-trivial moments are

$$M_2(\tau) = M_2(0)\, e^{-a_2 \tau},$$

$$M_3(\tau) = M_3(0)\, e^{-a_3 \tau},$$

$$M_4(\tau) = \left[M_4(0) + 3M_2^2(0) \right] e^{-a_4 \tau} - 3M_2^2(\tau).$$

To obtain explicit time dependence, we note that the second moment coincides with the granular temperature T. Thus we integrate $d\tau = \sqrt{T}\, dt$, with $T = T(0)\, e^{-a_2 \tau}$, to re-express the moment equations as

$$M_2(t) = T = \frac{T(0)}{(1 + a_2 t \sqrt{T(0)}/2)^2} \sim \left(\frac{2}{a_2 t} \right)^2,$$

$$M_3(t) = M_3(0)\, T^{a_3/a_2}, \tag{3.79}$$

$$M_4(t) = \left[M_4(0) + 3M_2^2(0) \right] T^{a_4/a_2} - 3T^2.$$

In stark contrast to equilibrium statistical mechanics, the temperature does not fully characterize the velocity distribution. For the classic Maxwell–Boltzmann distribution, $M_4 \sim T^2, M_6 \sim T^3$, and generally $M_{2k} \sim T^k$ – there is just one energy scale. However, this property does not hold in general for the inelastic Maxwell gas. From $a_n(\epsilon) = 1 - \epsilon^n - (1 - \epsilon)^n$, we have $a_3 = 3a_2/2$, so that asymptotic behavior of M_3 is given by $M_3 \sim M_2^{3/2}$. Thus the second moment continues to characterize the scaling behavior of the third moment. However, $a_4 < 2a_2$, so that the ratio M_4/M_2^2 diverges as $t \to \infty$. In general, the moments scale as

$$M_n \sim \begin{cases} T^{n/2}, & n \le n^* = 3, \\ T^{a_n/a_2}, & n \ge n^*, \end{cases} \tag{3.80}$$

where the a_n are independent coefficients for $n \ge n^*$. Since $T \to 0$ as $t \to \infty$, particles continuously slow down and eventually come to rest.

This change in the behavior of the moments indicates that the velocity distribution has a power-law, rather than a Gaussian, tail. To determine the velocity distribution, notice that the collision term in the Boltzmann equation (3.77) has a convolution structure, which suggests employing the Fourier transform. Defining $P(k, t) = \int e^{ikv} P(v, t)\, dv$ and computing the Fourier transform of the right-hand side of Eq. (3.77), we obtain the product form:

$$\iiint e^{ikv}\, P(u_1, \tau) P(u_2, \tau)\, \delta[v - \epsilon u_1 - (1 - \epsilon)u_2]\, du_1\, du_2\, dv$$

$$= \int e^{ik\epsilon u_1}\, P(u_1, \tau)\, du_1 \int e^{ik(1-\epsilon)u_2}\, P(u_2, \tau)\, du_2$$

$$= P(\epsilon k, \tau)\, P((1 - \epsilon)k, \tau).$$

Thus the Fourier transform of Eq. (3.77) is

$$\frac{\partial P(k,\tau)}{\partial \tau} + P(k,\tau) = P((1-\epsilon)k,\tau)\,P(\epsilon k,\tau)\,. \tag{3.81}$$

While this nonlinear and non-local equation appears intractable, we can make progress in the long-time limit, where the velocity distribution should have the scaling form $P(v,\tau) = T^{-1/2}\mathcal{P}(w)$, with scaling variable $w = v\,T^{-1/2}$. The corresponding scaling of the Fourier transform is $P(k,\tau) = \pi(z)$ with $z = |k|\,T^{1/2}$, and the two scaling functions are related, $\pi(z) = \int \mathcal{P}(w)\,e^{izw}\,dw$. Substituting this scaling ansatz into Eq. (3.81) and using the temperature decay rate $dT/d\tau = -2\epsilon(1-\epsilon)T$ from Eq. (3.78), the scaling function $\pi(z)$ satisfies

$$-\epsilon(1-\epsilon)z\,\frac{d\pi(z)}{dz} + \pi(z) = \pi(z - \epsilon z)\,\pi(\epsilon z)\,. \tag{3.82}$$

There does not appear to be a systematic way to solve Eq. (3.82), but by trial-and-error the solution is

$$\pi(z) = (1+z)\,e^{-z}\,, \tag{3.83}$$

which, surprisingly, is *independent* of the dissipation parameter. Inverting this Fourier transform, the scaled velocity distribution has the power-law form[20]

$$\mathcal{P}(w) = \frac{2}{\pi}\,\frac{1}{(1+w^2)^2} \rightarrow \frac{2}{\pi\,\sqrt{T}}\,\frac{1}{\left(1+v^2/T\right)^2}\,, \tag{3.84}$$

Unexpectedly, this distribution does not depend on the dissipation parameter ϵ. Moreover, by substituting Eq. (3.84) into the Boltzmann equation (3.77), we find that Eq. (3.84) is *exact* for *all* times. Thus an initial distribution of the form Eq. (3.84) is always a solution, while other initial conditions evolve to this scaling form only in the long-time limit.

For the more realistic case of the hard-sphere gas, the tail of the velocity distribution decays exponentially in velocity,

$$\mathcal{P}(w) \sim e^{-Cw}\,, \qquad w = \frac{\mathbf{v}}{\sqrt{T}}\,, \tag{3.85}$$

when $w \gg 1$. Heuristically, one can deduce this behavior by focusing only on the high-velocity tail of the distribution. At large velocities, the loss term is the dominant factor in the BE. In this limit, we may also replace $|\mathbf{v} - \mathbf{v}'|$ in the collision integral by $|\mathbf{v}|$. Solving the resulting differential equation $dP/dt \sim -|\mathbf{v}|P$ immediately leads to the exponential tail. This argument is hard to make rigorous, yet the prediction (3.85) appears to be correct. However, the solution of the BE for inelastic hard spheres still remains an outstanding open problem.

---

[20] The inverse Fourier transform of $e^{-\kappa z}$ is $(1/\pi)\kappa/(\kappa^2 + w^2)$; the inverse transforms of $z^n e^{-\kappa z}$ can be obtained using successive differentiation with respect to κ.

Driven gas

When energy is continuously injected into the system, the inelastic gas reaches a steady state where the energy input balances the loss due to inelastic collisions. Such an input can be realized by putting an inelastic gas in a closed container and shaking it. Experiments of this kind show that the velocity distribution generically has a non-Maxwellian tail for many geometries and energy injection mechanisms. A natural way to model this input is to impose a random acceleration on each particle, so its motion is described by the Langevin equation[21]

$$\frac{dv_\alpha}{dt} = \xi_\alpha , \tag{3.86}$$

in which $\langle \xi_\alpha \rangle = 0$ and $\langle \xi_\alpha(t)\xi_\beta(t') \rangle = 2D\delta_{\alpha\beta}\, \delta\left(t - t'\right)$. Because this random acceleration acts homogeneously throughout the system, the connection between this type of input with experiments in which only a small fraction of the particles at the walls of the container are directly influenced by the energy input is unknown. Nevertheless, one might hope that the tail of the velocity distribution will be a universal characteristic that does not depend on the precise form of the energy input.

Because the forcing in Eq. (3.86) leads to diffusion in velocity space, the corresponding BE (3.77) should be modified as

$$\frac{\partial P(v,t)}{\partial t} \rightarrow \frac{\partial P(v,t)}{\partial t} - D\frac{\partial^2 P(v,t)}{\partial v^2} . \tag{3.87}$$

Before delving into the analysis of the resulting BE, we first give a heuristic, but general, argument for the form of the high-energy tail of the velocity distribution. For sufficiently large velocities, the gain term in the Boltzmann equation (Eq. (3.77), with the time derivative augmented as in (3.87)), is negligible.[22] The resulting equation for the steady-state distribution becomes

$$D\frac{d^2 P(v)}{dv^2} \simeq P(v) , \tag{3.88}$$

which yields an exponential high-energy tail, $P(v) \sim e^{-v/v^*}$. This same argument can be extended in a simple way to arbitrary collision rates. For a collision rate that is proportional to $|u_1 - u_2|^\delta$ for $|u_1 - u_2| \to \infty$, the right-hand side in (3.88) becomes $-|v|^\delta P_\infty$, leading to $P_\infty(v) \sim \exp(-|v|^\gamma)$ with $\gamma = 1 + \delta/2$. For the classical case of hard spheres (collision rate linear in relative velocity), we obtain $\gamma = 3/2$. (See also problem 3.18.)

Following the same steps that led to Eq. (3.81) in the case of the freely cooling gas, the Fourier transform of the steady-state velocity distribution obeys

$$P(k) = \frac{1}{1 + Dk^2}\, P\left((1 - \epsilon)k\right)\, P(\epsilon k). \tag{3.89}$$

---

[21] Here we use Greek indices to label different particles.
[22] A similar partition of gain and loss terms also occurs in aggregation kinetics (see Chapter 5), which gives an easy way to determine the asymptotic properties of the cluster mass distribution.

Here the term that involves Dk^2 originates from the velocity diffusion term in Eq. (3.87). One way to solve (3.89) is by repeatedly substituting the left-hand side into the right-hand side. The first step in this approach gives

$$P(k) = \frac{1}{1 + Dk^2} \times \frac{1}{1 + D((1 - \epsilon)k)^2} \times \frac{1}{1 + D(\epsilon k)^2}$$
$$\times P\left((1 - \epsilon)^2 k\right) \times [P((1 - \epsilon)\epsilon k)]^2 \times P(\epsilon^2 k).$$

Continuing this process and using the boundary conditions $P(0) = 1$ and $P'(0) = 0$ imposed by the conservation of the total particle number and the total momentum, the solution is

$$P(k) = \prod_{l \geq 0} \prod_{m=0}^{l} \left[1 + \epsilon^{2m}(1 - \epsilon)^{2(l-m)} Dk^2 \right]^{-\binom{l}{m}}. \tag{3.90}$$

In the complex k-plane this function has an infinite series of simple poles at $\pm i \left[\epsilon^{2m}(1 - \epsilon)^{2(l-m)} D \right]^{-1/2}$. The high-velocity tail of $P(v)$ is governed by the two poles that are closest to the origin at $k = \pm i/\sqrt{D}$. Because the poles are simple and a finite distance from the origin, the velocity distribution has a pure exponential tail

$$P(v) \sim v_*^{-1} \, e^{-|v|/v_*} \qquad \text{as} \quad |v| \to \infty, \tag{3.91}$$

with $v_* = \sqrt{D}$, just as in the freely cooling Maxwell gas. A surprising feature is that the dependence on the dissipation parameter arises only in the prefactor.

3.7 Ballistic agglomeration

In *ballistic agglomeration*, freely moving particles stick together irreversibly whenever the two aggregates collide. Such a model provides an idealized description of the large-scale accretion of matter to form astronomical objects and the merging of thermal plumes or vortices. As a result of repeated collisions, aggregates grow continuously, with mass and momentum conserved in each collision. This model represents a simple application of ideas from this chapter to understand the consequences of collision kinetics at macroscopic scales.

We schematically write the agglomeration process as

$$(m_1, \mathbf{p}_1) + (m_2, \mathbf{p}_2) \to (m_1 + m_2, \mathbf{p}_1 + \mathbf{p}_2),$$

where m_i is the mass of the ith aggregate and v_i its velocity. We assume that aggregates are liquid drops of radii proportional to $m^{1/3}$ in three dimensions, so that immediately after any collision the resulting aggregate is again spherical (Fig. 3.6). The initial system is sufficiently dilute so that three- and higher-body interactions can be neglected. The system starts with a random spatial distribution of particles, with their initial velocities and masses drawn from some prescribed distributions. What is the evolution of this agglomeration process?
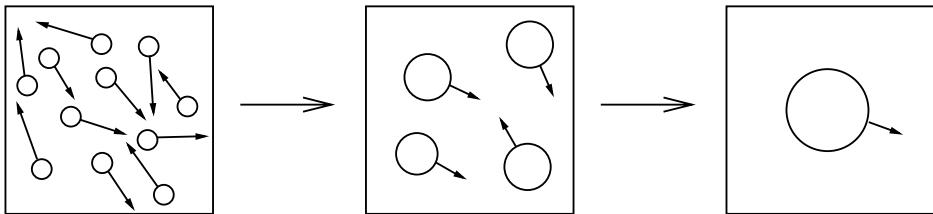
Fig. 3.6.   Schematic illustration of the evolution in ballistic agglomeration.

The time evolution of the typical cluster mass, speed, and density can be obtained by a back-of-the-envelope estimate. To simplify matters, suppose that all clusters initially have unit mass and unit velocities whose directions are random. An aggregate of mass m will therefore consist of m initial particles. If the initial particle momenta are uncorrelated, the momentum of an aggregate of mass m is the sum of m uncorrelated momenta of magnitude one, and therefore is of the order of $m^{1/2}$. Hence the typical velocity of such an aggregate is $v \sim p/m \sim m^{-1/2}$. To determine how the average mass grows with time, note that the mass of an aggregate increases by the order of its mass in a typical collision, and the typical time τ between collisions is related to the mean free path ℓ by $\ell \sim v\tau$. Thus the mass evolves in time as

$$\frac{dm}{dt} \sim \frac{m}{\tau} . \tag{3.92}$$

A collision occurs when the density times the volume of the collision tube swept out by the trajectory of an aggregate equals 1, namely, $n(v\tau)R^{d-1} \simeq 1$. Here n is the aggregate density and R is the typical aggregate radius. Rewriting these quantities in terms of the typical aggregate mass ($n \sim m^{-1}, v \sim m^{-1/2}, R \sim m^{1/d}$), we have

$$\tau \sim \frac{1}{nvR^{d-1}} \sim m \times m^{1/2} \times m^{-(d-1)/d} \sim m^{(d+2)/2d} . \tag{3.93}$$

Using this mass dependence for τ in (3.92), we obtain

$$m \sim t^{2d/(d+2)} , \qquad v \sim t^{-d/(d+2)} , \qquad n \sim t^{-2d/(d+2)} . \tag{3.94}$$

While the foregoing argument is appealing, it is erroneous when the spatial dimension is greater than one (and finite).[23] The error stems from the incorrect assumption that the momentum of an aggregate of mass m consists of the sum of m uncorrelated momenta. In fact, particles that move toward each other tend to collide more often, and vice versa for diverging particles. Thus aggregates that are smaller than the average mass consist of units whose momenta are more likely to be positively correlated, and vice versa for larger-than-average aggregates. In one dimension these correlations do not play a central role, because growing agglomerates necessarily sweep up a complete line segment of uncorrelated initial

---

[23] The first evidence that Eq. (3.94) is wrong came from simulations, which gave $m \sim t^\alpha$, with $\alpha \approx 0.86$ in two dimensions, rather than $\alpha = 1$ as predicted by (3.94).

velocities. However, correlations between the momenta of the constituent monomers in a single agglomerate surprisingly play an important role in greater than one dimension.

To help understand the role of correlations, it is useful to consider agglomeration in the *reaction-controlled limit*.[24] To construct this limit, we define a collision to cause cluster merging with probability δ, with δ vanishingly small, while a collision is elastic with probability $1 - \delta$. Because almost all collisions are elastic, the velocity distribution will be close to the Maxwell–Boltzmann form for each cluster, independent of its mass. The resulting reaction-controlled agglomeration is tractable and the result of a detailed (and lengthy) analysis of the Boltzmann equation for this model is

$$m \sim t^{2d/(d+3)}, \qquad v \sim t^{-(d+1)/(d+3)}, \qquad n \sim t^{-2d/(d+3)}, \qquad (3.95)$$

in excellent agreement with simulations of the reaction-controlled model. Since the heuristic argument that led to Eq. (3.94) equally applies generally, its failure in the reaction-controlled limit makes it plausible that the heuristic approach may also fail in the ballistically controlled limit.

Ballistic agglomeration can be solved exactly in one dimension for the joint mass–velocity distribution $P(m, v, t)$. This solution is tedious, and we merely summarize the main results:

- The typical cluster size grows as $t^{2/3}$, while the typical cluster speed decreases as $t^{-1/3}$, in agreement with the heuristic predictions of Eq. (3.94). The full mass-velocity distribution is given by

$$P(m, v, t) = t^{-1/3} \mathcal{P}(M, V), \qquad M = mt^{-2/3}, \quad V = vt^{1/3} . \qquad (3.96)$$

- The scaled distribution $\mathcal{P}(M, V)$ is computable, from which one can extract the separate velocity and mass distributions, $\Phi(V) = \int dM \, \mathcal{P}(M, V)$ and $\Psi(M) = \int dV \, \mathcal{P}(M, V)$. The two distributions approach the scaling forms $\Phi(V)$ and $\Psi(M)$, respectively.
- Factorization is violated, $\mathcal{P}(M, V) \neq \Phi(V) \Psi(M)$.
- The mass distribution has simple asymptotic behaviors:

$$\Psi(M) \sim \begin{cases} 1/\sqrt{M}, & M \to 0, \\ e^{-CM^3}, & M \to \infty. \end{cases} \qquad (3.97)$$

There is also an intriguing connection between ballistic agglomeration and the Burgers equation in one dimension (as mentioned briefly in Section 2.9 and to be discussed in more detail in Section 4.3). In this mapping, the average interparticle separation $\ell \sim t^{2/3}$ corresponds to the average distance between adjacent shocks, and the merging of agglomerates corresponds to the merging of localized wave packets.

---

[24] Chapter 13 on reaction kinetics will give more context for the distinction between transport-controlled and reaction-controlled limits.

To understand the evolution of a system such as ballistically agglomerating particles, molecular dynamics simulations are an extremely useful investigative tool. In its simplest version, molecular dynamics integrates the equation of motion for each particle over a suitably defined time interval. For ballistic agglomeration, this integration moves all particles along their own straight-line trajectories until a collision occurs. At each collision, the time until the next collision between each particle pair should be recomputed and the minimum of all these times defines the next collision. While this approach is straightforward, it is prohibitively inefficient. To determine the next collision directly requires of the order of N^2 operations in an N-particle system (N operations in one dimension), because the collision times between all pairs have to be computed to find the minimal such time.
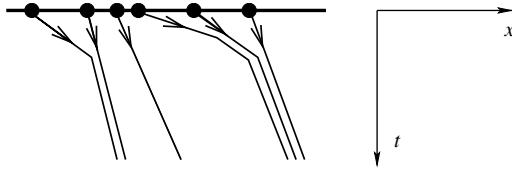
It is possible to substantially circumvent this inefficiency. For simplicity, we now specialize to the case of one dimension; however, the approach we now discuss can be largely adapted to higher dimensions. In one dimension, it is useful to initially sort the collision times of all nearest-neighbor pairs in increasing order by a standard $\mathcal{O}(N \ln N)$ sorting algorithm. Suppose that, according to this sorted list of collision times, particles n and $n + 1$ will agglomerate. To keep the list of collision times up to date, we must compute and store the new collision times that involve particles $n - 1$ and the newly agglomerated particle n, as well as between particles n and $n + 2$. An important feature of a more efficient algorithm is to continue to implement "virtual" agglomeration events in which particles are not moved but post-collision velocities continue to be recomputed by scanning through the precompiled list of collision times in ascending order. As soon as one of the newly created collision times becomes smaller than the next precompiled collision time, we propagate all particles over the sum of all ordered collision times up to the present. We also recompute and re-sort all collision times. Roughly speaking, we are able to look beyond the smallest collision time because the motion of a substantial fraction of the particles remains deterministic as the shortest-time collisions are implemented.

The moral here is that many of the inefficiencies of a straightforward molecular dynamics simulation can be substantially reduced by taking advantage of the feature that the motion of most of the particles remains deterministic for much longer than a single collision time.

3.8 Single-lane traffic

As many of us have experienced, traffic on rural highways, where there is a single lane available in each direction with no passing allowed, is inefficient. Faster cars accumulate behind slower-moving vehicles that catch up to vehicles that are moving more slowly still, so that almost no car is moving at its desired speed. The no-passing constraint causes initially homogeneous traffic to evolve to a clustered state. We can understand this frustrating phenomenon by kinetic theory.

We mimic traffic clustering by the following idealized model. Each car has its own intrinsic speed at which it would move on an empty road. When a faster car catches up to a

Schematic of traffic with no passing. Shown are the world-lines of cars for initially unclustered traffic (circles).

slower car, both cars subsequently move at the speed of the slower car (Fig. 3.7). Generally, when a cluster of m_2 cars, all moving at speed v_2, catches up to a cluster of m_1 cars that moves at speed $v_1 < v_2$, the resulting "collision" leads to a cluster of $m_1 + m_2$ cars that moves at speed v_1:

$$(m_1, v_1) + (m_2, v_2) \to (m_1 + m_2, \min(v_1, v_2)). \tag{3.98}$$

For simplicity, we assume (i) instantaneous collisions, (ii) sizeless cars, (iii) independent random initial speeds that are drawn from the same intrinsic distribution $P(v, 0)$, and (iv) initially uncorrelated initial car positions, with mean spacing equal to 1. Once the initial condition is specified, traffic evolves deterministically according to the no-passing collision rule (3.98). The only source of randomness is the initial conditions. Let us determine the evolution of this traffic flow.

Scaling

As a first step to understanding traffic clustering, we apply scaling to determine basic properties, such as the evolution of the typical cluster mass m, the separation between clusters ℓ, and their speed v. These quantities are controlled by the initial speed distribution of the slowest cars because asymptotically almost all cars accumulate in clusters behind these slow cars. As a generic example, we assume the initial speed distribution[25]

$$P(v, 0) \simeq a v^{\mu} \qquad \text{as} \quad v \to 0. \tag{3.99}$$

On dimensional grounds, the cluster separation ℓ and the speed v are related through $\ell \sim vt$. Because the separation between clusters should scale as the inverse of the concentration of clusters, which, in turn, is inversely proportional to the average number of cars in a cluster, we also have $\ell \sim m$. Now let's relate the number of cars in a cluster and its speed. Consider a car with speed v. The probability that a cluster of size k forms behind this car can be determined exactly for a "one-sided" problem in which cars exist only behind the target car. Let $P(v) = \int_0^v P(v', 0) \, dv'$ be the probability that the speed of a car is less than v. Then a cluster of size k eventually forms if there are $k - 1$ consecutive "fast" cars (cars with speed greater than v) followed by one slow car. The probability for

---

[25] Without loss of generality, we subtract the speed of the slowest car from all speeds so that the minimum speed is 0.

this configuration is $P(v)[1 - P(v)]^{k-1}$. The average size $\langle m(v) \rangle$ of the cluster behind a car of speed v is therefore

$$\langle m(v) \rangle = \sum_{k \geq 1} k\, P(v)[1 - P(v)]^{k-1} = \frac{1}{P(v)} \ .$$

This relation is exact for the one-sided problem; it is asymptotically exact in the small speed limit for the full two-sided problem because it is unlikely that a slow car will collide with even slower cars. Using the speed distribution (3.99) in $P(v)$, the average cluster size scales as $\langle m(v) \rangle \sim v^{-1-\mu}$. Combining this relation with $m \sim vt$ gives

$$m \sim t^{\alpha}, \qquad v \sim t^{-\beta} \qquad \text{with} \quad \alpha = \frac{\mu + 1}{\mu + 2}, \qquad \beta = \frac{1}{\mu + 2} \ . \qquad (3.100)$$

As one might anticipate, clusters of cars continuously grow with time while their average speed correspondingly decreases.

Exact speed distribution

We can proceed further and determine the speed distribution of free cars – cars that are not yet incorporated into a cluster. Let's begin with the special case where cars have only two possible speeds, v_1 and $v_2 > v_1$, with respective probabilities $P_1(0)$ and $P_2(0)$. Slow cars never catch any other vehicle, so their density is conserved, $P_1(t) = P_1(0)$. A fast car moves at its intrinsic speed v_2 before colliding with a slower car. To avoid a collision up to time t, a segment of length $(v_2 - v_1)t$ that is free of slow cars must exist ahead of a fast car. Since the initial spatial distribution of cars is random, this exclusion probability decays exponentially with the interval length $\exp[-P_1(0)(v_2 - v_1)t]$. Therefore, the density of the faster cars decays as

$$P_2(t) = P_2(0)\, e^{-P_1(0)(v_2 - v_1)t}. \qquad (3.101)$$

Now consider traffic with cars that move at three distinct intrinsic speeds $v_1 < v_2 < v_3$. Clusters with speeds v_1 and v_2 are unaffected by the presence of faster cars so the previous conclusions for their densities $P_1(t)$ and $P_2(t)$ continue to hold. On the other hand, for a fast car with speed v_3 to maintain its speed, it must avoid colliding with both clusters of speeds v_1 *and* v_2. The probability for these two independent events is given by a product of the two exclusion probabilities:

$$P_3(t) = P_3(0)\, e^{-P_1(0)(v_3 - v_1)t} \times e^{-P_2(0)(v_3 - v_2)t} \ . \qquad (3.102)$$

In general, when a car has an initial speed v_n with probability $P_n(0)$, for a (discrete) set of speeds $\{v_n\}$, with $v_1 < v_2 < v_3 < \ldots$, the cluster speed distribution is

$$P_n(t) = P_n(0) \prod_{1 \leq i \leq n-1} e^{-P_i(0)(v_n - v_i)y}$$

$$\to P(v, 0)\, \exp\left[-t \int_0^v dv'\, (v - v') P(v', 0) \right], \qquad (3.103)$$

where the second line is obtained by replacing the product by a sum in the exponential and then taking the continuum limit. Thus the density of cars of any speed decays exponentially in time, with a decay rate that is a growing function of v.

Using the initial speed distribution $P(v, 0) \simeq av^\mu$, Eq. (3.103) gives, in the long-time limit,

$$P(v, t) = a\, v^\mu\, \exp\left[-b\, v^{\mu+2}t\right], \qquad \text{with} \quad b = \frac{a}{(\mu + 1)(\mu + 2)}, \qquad (3.104)$$

which can be written in the scaling form

$$P(v, t) = t^{\beta-\alpha}\, \Phi(V), \qquad\qquad \Phi(V) = aV^\mu\, \exp\left[-bV^{\mu+2}\right] \qquad \text{with} \quad V = vt^\beta, \qquad (3.105)$$

with exponents $\alpha = (\mu + 1)/(\mu + 2)$ and $\beta = 1/(\mu + 2)$, as already predicted heuristically in (3.100). In accord with everyday intuition, the probability that a car with intrinsic speed v can drive freely without catching slower traffic quickly decays with time and is also a rapidly decreasing function of v – the faster the intrinsic speed of a car, the more quickly it will become stuck behind a bus.

This idealized traffic model nicely illustrates the shortcoming of the Boltzmann approach for one-dimensional systems. We can express Eq. (3.103) in a Boltzmann-like form by differentiating the logarithm of this equation with respect to time to give the *linear* evolution equation

$$\frac{\partial P(v, t)}{\partial t} = -P(v, t)\int_0^v dv'(v - v')P(v', 0). \qquad (3.106)$$

This equation resembles the classic Boltzmann equation (3.6), except that there does not exist a gain term because a cluster of a given speed cannot be created if it does not already exist. However, there are fundamental differences that are worth emphasizing:

- Equation (3.106) is *linear* in the velocity distribution $P(v, t)$. The classical Boltzmann equation is nonlinear.
- There is perpetual memory, as the initial condition $P(v, 0)$ continues to affect clustering *ad infinitum*.
- The Boltzmann equation for the single-lane traffic model is

$$\frac{\partial P(v, t)}{\partial t} = -P(v, t)\int_0^v dv'(v - v')P(v', t). \qquad (3.107)$$

In contrast to the linear (and exact!) equation (3.106), which is solvable for an arbitrary initial condition, the nonlinear Boltzmann equation (3.107) has not been solved for any initial condition.

- Because the Boltzmann equation (3.107) is based on a mean-field description of interparticle interactions, it should be expected to fail in one dimension, where only nearest-neighbor interaction occur. Equation (3.107) is indeed an uncontrolled approximation.

- The classical Boltzmann equation is based on factorizing many-particle correlation functions into a product of single-particle correlation functions. For example, the probability $P_2(v, w; t)$ that two cars have speeds v and w, with $w > v$, would be written as $P(w, t)P(v, t)$ in the Boltzmann approach. However, the correct result is $P_2(v, w; t) = P(w, t)P(v, 0)$, since the leading faster car "shields" the slower car behind and it therefore suffices to guarantee that the leading car has not been slowed down. This self-evident lack of factorization illustrates yet another shortcoming of the Boltzmann approach.

Many other correlation functions in the traffic model also do not factorize. Perhaps the most basic is the joint size–speed distribution, namely, the density of clusters that consist of m cars that move at speed v, $P_m(v, t)$. Scaling alone predicts that

$$P_m(v, t) = t^{\beta - \alpha} \, \mathcal{P}(M, V), \qquad M = mt^{-\alpha}, \quad V = vt^{\beta} \,. \tag{3.108}$$

The lack of factorization means that the scaling function $\mathcal{P}(M, V)$ cannot be written as a product $\Phi(V) \, \Psi(M)$. Strictly speaking, the non-factorizability is not a manifestation of the failure of the Boltzmann approach; the latter applies to many interacting entities, while $\mathcal{P}(M, V)$ describes a single-cluster entity. However, the lack of factorizability appears to be inherent in basic observables of traffic flow, such as $\mathcal{P}(M, V)$.[26]

3.9 Notes

Collision theory is the foundation upon which non-equilibrium statistical physics has been built. The symmetry argument used to derive the velocity distribution of a gas was given by Maxwell himself [31]. For the fundamentals of kinetic theory and the Boltzmann equation, there are the lecture notes of Boltzmann [32], as well as various books [33, 34] and review articles [35–37]. The derivation of hydrodynamics from the Boltzmann equation is a standard subject that can be found in almost every text on the Boltzmann equation, see e.g. [35, 38].

Drude's paper is [39] and the proposal for the Lorentz model is given in [40]. For a popularized account of the Lorentz model, see [41], while a comprehensive discussion of basic properties of this model is given in [42, 43]. The non-stationarity of the field-driven Lorentz gas and its physical consequences can be found in [44]. The text by Kubo *et al.* [45] presents the Green–Kubo formalism that we used to determine the diffusion coefficient of the Lorentz gas; this Green–Kubo formalism can also be used to obtain many useful fluctuation–dissipation relations. The impact problem and the closely analogous propagation of a continuum blast wave are treated in [9, 46–48]. The Maxwell model is discussed in [31, 35, 36]; more information about the very hard-particle model can be found in [36].

The unusual properties of inelastic gases, namely large-scale clustering and inelastic collapse, became popular in the 1990s. One of the early works on this subject is that of

---

[26] As we mentioned earlier, the scaled distribution $\mathcal{P}(M, V)$ that describes the agglomeration process in one dimension also does not factorize.

Haff [49], in which the cooling rate of an inelastic gas was elucidated. Figure 3.4 that shows inelastic collapse is taken from the article by Goldhirsch and Zanetti [50]. The exact solution for the free cooling of the inelastic Maxwell model was derived in [51]. The kinetic theory of inelastic gases is comprehensively discussed in the monograph by Brilliantov and Pöschel [52]. The Fourier transform technique for Maxwell molecules greatly simplifies the governing equations. This insight appears in the unpublished theses by Krupp [53]. More details about this approach in the context of elastic Maxwell molecules can be found in [36, 54]; the inelastic Maxwell model is reviewed in [55].

Ballistic agglomeration [56–59] is a subject that continues to evolve [60], and a comprehensive understanding has been achieved only in one dimension [56, 59, 61]. Much more about molecular dynamics simulation techniques and general algorithms for sorting and search through lists can be found, for example, in [62, 63]. The application of collision theory to one-dimensional traffic is based on [64].

3.10 Problems

3.1 The Boltzmann equation can be applied when a gas is dilute, namely when $nd^3 \ll 1$, where n is the density and d is the range of the intermolecular force. Under standard conditions (room temperature and atmospheric pressure), estimate the density of the atmosphere and also the range of intermolecular forces. Show that under these conditions the atmosphere is indeed dilute.

3.2 Justify that quantum effects are negligible for the atmosphere under standard conditions. In particular, show that the de Broglie wavelength is much smaller than the average distance between adjacent particles:

$$\frac{\hbar}{\sqrt{2mT}} \ll n^{-1/3} .$$

For a gas at atmospheric pressure, estimate the temperature at which quantum effects become important. For a gas at room temperature, estimate the density at which quantum effects become important.

3.3 Consider the scattering of a uniform parallel beam of point particles that moves in the z direction on an immobile spherical obstacle in d dimensions. Show that the ratio of the total backward scattering flux Φ_b to forward scattering flux Φ_f is

$$\frac{\Phi_b}{\Phi_f} = \frac{\left(1/\sqrt{2}\right)^{d-1}}{1 - \left(1/\sqrt{2}\right)^{d-1}} \equiv \lambda_d.$$

Here forward flux refers to particles that have a positive z component (and vice versa for the backward flux). In particular $\lambda_1 = \infty$ (only backward scattering in one dimension), $\lambda_2 = (\sqrt{2} - 1)^{-1} > 1$ (predominantly backward scattering in two dimensions), and $\lambda_3 = 1$ (a manifestation of isotropic scattering in three dimensions).

3.4 This problem is concerned with the derivation of the Einstein–Green–Kubo formula.

    (a) Show that the diffusion equation implies that the mean-square displacement $\Delta(t) = \langle [\mathbf{r}(t) - \mathbf{r}(0)]^2 \rangle$ grows as $\Delta \simeq 2dDt$.

    (b) Derive from the above result

$$\frac{d\Delta}{dt} = 2\langle \mathbf{v}(t) \cdot [\mathbf{r}(t) - \mathbf{r}(0)] \rangle = 2 \int_0^t dt' \, \langle \mathbf{v}(t) \cdot \mathbf{v}(t') \rangle .$$

    (c) Argue that the equilibrium average depends only on the difference of the time arguments $t'' = t - t'$ and therefore

$$\frac{d\Delta}{dt} = 2 \int_0^t dt'' \, \langle \mathbf{v}(0) \cdot \mathbf{v}(t'') \rangle .$$

    (d) Using $\Delta \simeq 2dDt$ and taking the $t \to \infty$ limit show that the previous equation yields

$$D = \frac{1}{d} \int_0^\infty dt \, \langle \mathbf{v}(0) \cdot \mathbf{v}(t) \rangle .$$

3.5 *Any* isotropic velocity distribution is an equilibrium distribution for the Lorentz model. If, however, there is a weak interaction between moving particles, they will eventually equilibrate at some temperature T so that the velocity distribution is Maxwellian. Show that in this situation the average diffusion coefficient (3.9) becomes

$$D = \frac{4\ell}{3} \sqrt{\frac{T}{2\pi m}} = \frac{16a}{9} \sqrt{\frac{T}{2\pi m}} \frac{1}{\nu} ,$$

    where $\nu = n(4\pi a^3/3) \ll 1$ is the volume fraction occupied by the scatterers.

3.6 Verify that Eq. (3.35) is the solution to the telegraph equation.

3.7 Using (3.47a) compute the root mean-square velocity of a Lorentz gas particle in an electric field. Show that this velocity coincides with the estimate $(\mathcal{E}^2 \ell t)^{1/3}$ up to a numerical factor $3^{2/3}[\Gamma(1/3)]^{-1/2}$.

3.8 Using the Green–Kubo formula as in the case of zero magnetic field calculate the non-trivial components of the diffusion tensor in Eq. (3.52) and show that they are given by

$$D_\parallel = \frac{1}{3} v_0^2 \tau, \quad D_\perp = \frac{1}{3} v_0^2 \tau \frac{1}{1 + (\omega\tau)^2} , \quad D_{xy} = -D_{yx} = -\frac{1}{3} v_0^2 \tau \frac{\omega\tau}{1 + (\omega\tau)^2} .$$

3.9 Use a kinetic theory argument, similar to that developed for the number of moving particles, Eq. (3.55), to determine the number of collisions $C(t)$ up to time t in the collision cascade discussed in Section 3.4.

3.10 Consider a zero-temperature gas that occupies the half-space $x \geq 0$. When a particle that moves along the x axis hits this system, it generates a collisional cascade that propagates into the medium. Particles are also ejected backward to create a splash-like pattern (Fig. 3.3). The goal of this problem is to argue that the energy $E(t)$ of the

particles in the initially occupied half-space ultimately approaches zero algebraically in time.

(a) Using dimensional analysis show that the size R of the cascade is given by Eq. (3.56), although we should now think about the energy $E(t)$ decreasing with time.

(b) Show that the total number of particles in the cascade scales as

$$N \sim nR^d \sim (E/m)^{d/(d+2)} \left(nt^d\right)^{2/(d+2)} .$$

(c) Show that the typical velocity of the ejected particles v is related to N and E via $Nmv^2 \sim E$, from which

$$v \sim \left(\frac{E}{nmt^d}\right)^{1/(d+2)} .$$

(d) Argue that the energy $E(t)$ decreases according to

$$\frac{dE}{dt} \sim -\frac{mv^2}{2} \times nvR^{d-1} .$$

(e) Using results from (a)–(c) show that (d) reduces to $dE/dt \sim -E/t$, which leads to an algebraic decay of E with time.

3.11 This problem concerns heat conduction in a medium where the heat conduction coefficient depends algebraically on temperature: $\kappa = \kappa(T) = AT^n$. (For Maxwell molecules, hard spheres, and very hard particles, the exponent is $n = 1, 1/2, 0$, respectively, following the argument that led to (3.57) for the diffusion coefficient.) The heat conduction equation

$$\frac{\partial T}{\partial t} = A\nabla \cdot (T^n \nabla T)$$

is mathematically a *parabolic* partial differential equation. When $n = 0$, there is a general solution for an arbitrary initial condition. When $n \neq 0$, the resulting nonlinear parabolic equation is generally insoluble, but exact solutions are known for sufficiently simple initial conditions. Consider one such situation where heat is instantaneously released on the (hyper)plane $x = 0$ while outside the initial temperature is zero. Then we need to solve the initial value problem

$$\frac{\partial T}{\partial t} = A\frac{\partial}{\partial x}\left(T^n\frac{\partial T}{\partial x}\right), \qquad T(x,0) = Q\delta(x).$$

The goal is to solve this initial value problem when $n \geq 0$.

(a) On general grounds, $T = T(x, t\,|A, Q)$. Using dimensional analysis show that any three out of four quantities x, t, A, Q have independent dimensions.

(b) Choosing t, A, Q as the basic quantities, show that the solution must have the form

$$T(x,t) = \left(\frac{Q^2}{At}\right)^{1/(2+n)} f(\xi), \qquad \xi = \frac{x}{(Q^n At)^{1/(2+n)}} .$$

Thus dimensional analysis alone shows that the solution (which in principle depends on two variables x and t) is, up to the scaling factor, a function of a single scaling variable ξ. The primary reason for this simplification is the absence of a quantity with dimensions of length in the formulation of the initial value problem.

(c) Substitute the above ansatz for $T(x, t)$ into the nonlinear heat equation and show that the scaling function satisfies the ordinary differential equation

$$(f^n f')' + \frac{1}{2+n} (\xi f)' = 0$$

where the prime denotes differentiation with respect to ξ.

(d) Show that the solution to the above differential equation is

$$f(\xi) = \left[\frac{n}{2(2+n)} (\xi_0^2 - \xi^2) \right]^{1/n}$$

in the region $|\xi| \le \xi_0$ and $f(\xi) = 0$ for $|\xi| \ge \xi_0$. Argue that ξ_0 is determined by heat conservation

$$Q = \int T(x, t)\, dx = Q \int_{-\xi_0}^{\xi_0} f(\xi)\, d\xi$$

and compute ξ_0.

(e) The heated region is $[-x_0(t), x_0(t)]$ whose boundaries grow as $x_0(t) = \xi_0 (Q^n At)^{1/(2+n)}$; outside the heated region the medium remains at zero temperature. Try to explain the contradiction with the common wisdom that, for a parabolic partial differential equation, perturbations propagate instantaneously.

3.12 Consider the previous problem in three dimensions and assume that a finite amount of heat has been instantaneously released at the origin: $T(\mathbf{r}, 0) = Q\delta(\mathbf{r})$.

(a) Using the same arguments as in the previous problem show that the solution must have the form

$$T(\mathbf{r}, t) = \left(\frac{Q^2}{A^3 t^3} \right)^{1/(2+3n)} f(\xi), \qquad \xi = \frac{r}{(Q^n At)^{1/(2+3n)}}.$$

(b) Derive an ordinary differential equation for $f(\xi)$ and solve it to give

$$f(\xi) = \left[\frac{n}{2(2+3n)} (\xi_0^2 - \xi^2) \right]^{1/n}.$$

(c) The heated region is a ball of radius $r_0(t) = \xi_0 (Q^n At)^{1/(2+3n)}$. Thus the volume of the heated region grows as $t^{3/5}$ and $t^{6/7}$ in the case of Maxwell molecules and hard spheres, respectively; for very hard spheres, the entire space $\mathbb{R}^3$ becomes heated at any $t > 0$.

3.13 Consider heat conduction in the three-dimensional region $R \le r < \infty$. Assume that the inner spherical boundary is maintained at a constant temperature, $T(R, t) = T_0$,

and that the temperature vanishes at infinity, $T(\infty, t) = 0$. Show that, if $\kappa(T) = AT^n$, the steady-state temperature satisfies $\nabla^2(T^{n+1}) = 0$, and that the corresponding temperature profile is $T(r) = T_0 (R/r)^{1/(n+1)}$.

3.14 Consider ballistic annihilation in one dimension. Transform the governing integro-differential Boltzmann equation (3.59) into

$$\frac{\partial}{\partial t} \frac{\partial^2}{\partial v^2} \ln P(v, t) = -2P(v, t).$$

This (apparently very difficult) nonlinear partial differential equation has not been solved so far.

3.15 Explain the relation $\alpha + \beta = 1$ between the decay exponents that appear in (3.60).

3.16 Consider ballistic annihilation for very hard particles. Confirm the expressions (3.68) for the density and the mean speed and show that the decay exponent β is indeed given by (3.69).

3.17 Verify the general behavior of the moments (quoted in Eq. (3.80)) for the one-dimensional inelastic Maxwell gas by exploiting the inequality $a_n < a_m + a_{n-m}$ for all $1 < m < n - 1$.

3.18 Suppose that the dependence of the collision rate on relative velocity is $|u_1 - u_2|^\delta$ when $|u_1 - u_2| \to \infty$. Use this fact to determine the analog of Eq. (3.88), as well as the asymptotic form of $P(v, t)$ for large v.

3.19 Consider the following toy ballistically controlled aggregation model in which the masses of all particles always remain identical. That is, when two particles with velocities $\mathbf{v}_1$ and $\mathbf{v}_2$ collide, they form an aggregate with velocity $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2$.

(a) Show that for the toy model the volume fraction occupied by particles decays indefinitely.

(b) Does the Boltzmann framework provide an asymptotically exact description? From (a) we see that the system becomes dilute, so that multiple collisions become asymptotically irrelevant.[27] What do you think about the behavior in one dimension? (The molecular chaos assumption, and generally a mean-field approach, is always suspicious in one dimension.)

(c) Using a (simplified) version of the mean-field argument applied to establish the conjectural behavior for ballistic agglomeration, Eqs (3.92)–(3.93), show that for the toy model the analog of (3.94) reads

$$v \sim t, \quad n \sim t^{-2}$$

independent of the spatial dimension d.

(d) Show that the Boltzmann equation describing the toy model is

$$\frac{\partial P(\mathbf{v}, t)}{\partial t} = \int d\mathbf{u}\, d\mathbf{w}\, P(\mathbf{u}, t)\, P(\mathbf{w}, t)\, |\mathbf{u} - \mathbf{w}|\, \delta(\mathbf{u} + \mathbf{w} - \mathbf{v})$$

$$- 2P(\mathbf{v}, t) \int d\mathbf{w}\, P(\mathbf{w}, t)\, |\mathbf{v} - \mathbf{w}|.$$

---

[27] For the ballistic agglomeration model, the volume fraction remains constant, so the Boltzmann equation approach can apply only if the system is initially dilute.

(Numerical solution of this BE shows that the density decays algebraically, $n \sim t^{-\alpha}$, with exponent increasing with dimension: $\alpha \approx 1.33, 1.55, 1.65$ when $d = 1, 2, 3$.)

(e) Simulate the BE in one dimension and show that the exponent is indeed $\alpha \approx 4/3$, very different from $\alpha = 2$ implied by the mean-field argument.

(f) Consider the toy model in the reaction-controlled limit. Show that $n \sim t^{-2d/(d+1)}$. The mean-field argument does not distinguish the ballistically controlled limit from the reaction-controlled one, so its prediction $n \sim t^{-2}$ is wrong in any dimension.

3.20 Assume that a cluster consists of m cars with velocities independently drawn from the distribution $P_0(v)$. Using the asymptotic form $P_0(v) \sim v^\mu$ as $v \to 0$, show that $m \sim v^{-\mu-1}$.

3.21 Evaluate the cluster size distribution $P(v, t)$ for the special case $P_0(v) = e^{-v}$. Show that the result is consistent with the general scaling behavior.

3.22 Consider the simplest heterogeneous initial condition: a half-line $x < 0$ is occupied by cars whose positions and velocities are uncorrelated, while the other half-line $x > 0$ is initially empty. Show that the total number of clusters that infiltrate the initially empty half-line grows as $\alpha \ln t$. Here $\alpha = (\mu + 1)/(\mu + 2)$ is the exponent that characterizes the number of cars in the typical cluster (in the bulk) $m \sim t^\alpha$.

Exclusion

The goal of statistical physics is to study collective behaviors of interacting many-particle systems. In equilibrium statistical physics, the simplest interaction is exclusion – for example, hard spheres that cannot overlap. This model depends on a single dimensionless parameter, the volume fraction; the temperature is irrelevant since the interaction energy is zero when the spheres are non-overlapping and infinite otherwise. Despite its apparent simplicity, the hard-sphere gas is incompletely understood except in one dimension. A similar state of affairs holds for the lattice version of hard spheres; there is little analytical understanding of its unusual liquid–gas transition when the spatial dimension $d \geq 2$.

In this chapter we explore the role of exclusion on the simplest *non-equilibrium* models that are known as *exclusion processes*. Here particles occupy single lattice sites, and each particle can hop to a neighboring site only if it is vacant (see Fig. 4.1). There are many basic questions we can ask: What is the displacement of a single particle? How does the density affect transport properties? How do density gradients evolve with time? In greater than one dimension, exclusion does not change transport properties in any qualitative way compared to a system of independent particles. However, exclusion leads to fundamentally new transport phenomena in one dimension.

4.1 Symmetric exclusion process

In the symmetric exclusion process (SEP), the average motion of each particle is spatially symmetric. For a single particle, the SEP is identical to the random walk. In the complementary limit where the lattice is fully occupied except for a single empty site, this vacancy
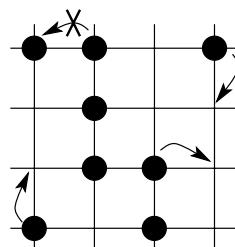

Fig. 4.1. Illustration of the symmetric exclusion process on the square lattice. A particle can move to a neighboring site if it is vacant.

also undergoes a simple random walk. A surprising fact, at first sight, is that, for *any* particle concentration $0 \leq c \leq 1$, the overall density in the SEP evolves as if there is no exclusion.

Evolution of the density

Let $P_n(t)$ be the single-particle density, namely the probability that a particle occupies lattice site n at time t. We also need two different pair densities: $P_{n,m}(t)$, the joint probability that sites n and m are occupied, and $P_{n,\overline{m}}(t)$, the joint probability that n is occupied and m is vacant. For the SEP, the single-particle density evolves by

$$\frac{dP_n}{dt} = \sum_m \left(P_{\overline{n},m} - P_{n,\overline{m}} \right),$$ (4.1)

where the sum runs over all sites m that are neighbors of n. Namely, a particle can hop to site n if this site is vacant and one its neighbors m is occupied, while a particle can leave n if one of its neighbors is vacant.

We can eliminate the joint probabilities in favor of single-particle densities by noting that the joint probabilities satisfy the obvious identities:

$$P_{n,m}(t) + P_{n,\overline{m}}(t) = P_n(t),$$

$$P_{n,m}(t) + P_{\overline{n},m}(t) = P_m(t),$$

as well as the symmetry condition $P_{n,m}(t) = P_{m,n}(t)$. Using these facts, we may eliminate $P_{\overline{n},m}$ and $P_{n,\overline{m}}$ in Eq. (4.1) and immediately obtain

$$\frac{dP_n}{dt} = \sum_m \left(P_m - P_n \right).$$ (4.2)

This is just the master equation for random walk hopping of a single particle! This equivalence holds for any lattice. Thus exclusion has no effect on the spatial and temporal evolution of the density. In the large-distance and long-time limit, the density is therefore governed by the classical diffusion equation.

Evolution of a single tracer particle

Despite the similarity with classical diffusion, exclusion does play an essential role in one dimension when we focus on the motion of one specific particle – a *tagged* or, equivalently, a *tracer* particle. Clearly, a tracer cannot wander as far as in free diffusion. When the spatial dimension d is 2 or greater, this hindrance manifests itself as a mere renormalization of the diffusion coefficient: $\langle \mathbf{r}^2 \rangle = 2d\, D(c)\, t$, with a concentration-dependent diffusion coefficient $D(c)$ that decreases with c. However, in one dimension, exclusion leads to a root-mean-square displacement that grows as

$$x_{\text{rms}} = \sqrt{\langle x^2 \rangle} \simeq A\, t^{1/4} \qquad \text{with} \quad A = \left(\frac{2}{\sqrt{\pi}} \frac{1-c}{c} \right)^{1/2}.$$ (4.3)

In spite of this anomalous time dependence, the probability distribution for the position of the tracer is still a Gaussian function, but now in the variable $x/t^{1/4}$.

We can intuitively understand the time dependence of the tracer displacement by considering the nearly full lattice, $c \to 1$. In this limit, any particle is almost always surrounded by occupied sites and therefore rarely moves. Conversely, since the vacancy density $1 - c$ is small, vacancies barely "see" each other and diffuse freely. The key point is that the motion of the tracer can be inferred from the motion of these freely diffusing vacancies. A tracer will hop to the right when its right neighbor contains a vacancy, and similarly for hopping to the left. Thus the displacement of a tracer is

$$x(t) = N_+(t) - N_-(t), \tag{4.4}$$
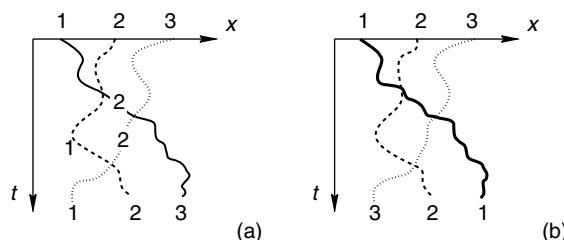
where $N_+(t)$ is the number of vacancies that were initially to the right of the tracer and are now on the left, and vice versa for $N_-(t)$. Since vacancies diffuse freely and their densities to the right and to the left of the tracer are equal, the number of trajectory crossings N_+ and N_- are given by

$$\langle N_+(t) \rangle = \langle N_-(t) \rangle \sim t^{1/2}. \tag{4.5}$$

Thus the average tracer displacement $\langle x \rangle$ is zero. However, the random-walk-like difference in the number of steps to the left and right scales as $\sqrt{t^{1/2}}$ and thus the mean-square displacement grows as $t^{1/2}$, in agreement with (4.3).

To derive the probability distribution $P(x, t)$ for tracer displacements, it is much simpler to consider a continuum variant of the SEP in which background particles diffuse independently and pass freely through any other particle whenever an encounter occurs (Fig. 4.2). In contrast, when two particles meet in the SEP, they collide and reflect so that their order is preserved. Crucially, both interaction rules lead to the *same* evolution for the tracer. Consequently, we deal with the simpler situation of independent and freely diffusing background particles to determine the displacement distribution of a tracer particle.

Suppose that the tracer starts at the origin and reaches x at time t. Then the average number of freely diffusing background particles that were initially to the right of the tracer


Fig. 4.2. Equivalence of free diffusion and the exclusion process. Shown are the world-lines of three particles in one dimension. (a) With exclusion, particles collide when they meet so that their labels are always ordered. (b) For free diffusion, particle labels remain attached to a specific trajectory (solid, dashed, dotted).
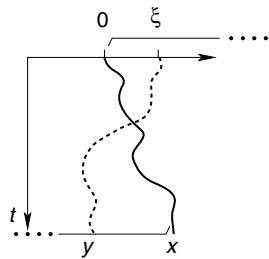
Fig. 4.3. **Illustration of Eq. (4.6a). A background particle initially at ξ to the right of the tracer has crossed the tracer by time t and reached $y < x$.**

and are now to its left at time t is (Fig. 4.3)

$$\nu_+(x,t) \equiv \langle N_+(x,t) \rangle = \int_0^\infty d\xi \int_{-\infty}^x \frac{\rho}{\sqrt{4\pi Dt}} \, e^{-(y-\xi)^2/4Dt} \, dy \, . \qquad (4.6a)$$

Here ξ is the initial position of a background particle, y is its final position, and ρ is the background density. Similarly

$$\nu_-(x,t) \equiv \langle N_-(x,t) \rangle = \int_{-\infty}^0 d\xi \int_x^\infty dy \, \frac{\rho}{\sqrt{4\pi Dt}} \, e^{-(y-\xi)^2/4Dt} \qquad (4.6b)$$

is the average number of background particles that were initially to the left of the tracer and are to its right at time t. Obviously $\nu_+(-x,t) = \nu_-(x,t)$.

What is the probability $P_+(n,x,t)$ that n background particles that were initially to the right of the tracer are on its left at time t? Since these particles move independently, this probability is Poissonian with average value $\nu_+(x,t)$, and similarly for $P_-(n,x,t)$. Therefore

$$P_\pm(n,x,t) = \frac{\nu_\pm^n}{n!} \, e^{-\nu_\pm} \, . \qquad (4.7)$$

The probability $\Pi_k(x,t)$ that there are $k \geq 0$ more crossings from the right than from the left (the case of $k < 0$ is similar) is then

$$\Pi_k(x,t) = \sum_{n\geq 0} P_+(n+k,x,t) P_-(n,x,t) \, . \qquad (4.8)$$

Substituting the expressions for $P_\pm(n,x,t)$ from (4.7) into the above sum, and using the series expansion for the Bessel function from the highlight below, we immediately obtain

$$\Pi_k(x,t) = e^{-(\nu_+ + \nu_-)} \left(\frac{\nu_+}{\nu_-} \right)^{k/2} I_k\big(2\sqrt{\nu_+ \nu_-}\big) \, . \qquad (4.9)$$

A standard definition of the modified Bessel function that applies for arbitrary real index $\mu \geq 0$ is via the differential equation

$$z^2 \frac{d^2 y}{dz^2} + z \frac{dy}{dz} - (z^2 + \mu^2)y = 0. \tag{4.10}$$

This second-order equation has two linearly independent solutions, $I_\mu(z)$ and $K_\mu(z)$, the modified Bessel functions of the first and second kind, respectively. The modified Bessel function of the first kind admits the series representation for $z \to 0$:

$$I_\mu(z) = \left(\frac{z}{2}\right)^\mu \sum_{m \geq 0} \frac{1}{m!\,(m+\mu)!} \left(\frac{z}{2}\right)^{2m}, \tag{4.11}$$

while the modified Bessel function of the second kind $K_\mu(z)$ diverges as $z \to 0$.

The crucial insight is that $P(x,t)$ is directly related to $\Pi_0(x,t)$. As illustrated in Fig. 4.2, if we denote particle 2 as the tracer, its position at time t in the exclusion process (Fig. 4.2(a)) coincides with the position of a freely diffusing particle that experiences an *equal* number of crossings from the left and the right by the background particles (Fig. 4.2(b)). Thus the distributions $P(x,t)$ and $\Pi_0(x,t)$ are equivalent.

We now determine the precise relation between $\Pi_0(x,t)$ and the tracer probability distribution $P(x,t)$ in the long-time limit. We will assume that the relevant length scale is smaller than diffusive, $|x| \ll t^{1/2}$, and the final result will show that this assumption is self-consistent. We start by rewriting the second integral in Eq. (4.6a) as

$$\int_{-\infty}^{x} \frac{\rho}{\sqrt{4\pi Dt}}\, e^{-(y-\xi)^2/4Dt}\, dy = \frac{\rho}{\sqrt{\pi}} \int_{-\infty}^{(x-\xi)/\sqrt{4Dt}} e^{-z^2}\, dz\,,$$

and then interchanging the integration order in (4.6a) to simplify ν_+ to

$$\nu_+ = \frac{\rho}{\sqrt{\pi}} \int_0^\infty d\xi \int_{-\infty}^{(x-\xi)/\sqrt{4Dt}} e^{-z^2}\, dz$$

$$= \frac{\rho}{\sqrt{\pi}} \int_{-\infty}^{x/\sqrt{4Dt}} e^{-z^2}\, dz \int_0^{x - z\sqrt{4Dt}} d\xi$$

$$= \rho\sqrt{4Dt/\pi} \int_{-\infty}^{\eta} (\eta - z)\, e^{-z^2}\, dz\,,$$

where $\eta = x/\sqrt{4Dt}$. Similarly

$$\nu_- = \rho\sqrt{4Dt/\pi} \int_\eta^\infty (z - \eta)\, e^{-z^2}\, dz\,.$$

Our assumption $|x| \ll t^{1/2}$ is equivalent to $|\eta| \ll 1$. In this limit

$$v_\pm(\eta, t) = \rho\sqrt{Dt/\pi}\left(1 \pm \eta\sqrt{\pi} + \eta^2 + \cdots\right). \tag{4.12}$$

We also use the long-time asymptotic form of the Bessel function, $I_0(2t) \simeq e^{2t}/\sqrt{4\pi t}$, to recast (4.9) for $k = 0$ as

$$\Pi_0(x, t) \simeq (4\pi)^{-1/2}(v_+ v_-)^{-1/4}\, e^{-(\sqrt{v_+} - \sqrt{v_-})^2}.$$

Substituting (4.12) into the above result and keeping only the dominant terms, we obtain the probability for an equal number of left- and right-crossings of the tracer:

$$\Pi_0(x, t) \simeq \frac{1}{\sqrt{4\pi\rho}}\left(\frac{\pi}{Dt}\right)^{1/4}\, e^{-\rho x^2/4\sqrt{Dt/\pi}}.$$

Integrating this expression over all x gives $1/\rho$. Hence the probability distribution of displacements of the tracer is indeed the Gaussian

$$P(x, t) = \rho\, \Pi_0(x, t) = \frac{1}{\sqrt{2\langle x^2\rangle}}\, e^{-x^2/2\langle x^2\rangle} \qquad \text{with} \quad \langle x^2\rangle = \frac{2}{\rho}\sqrt{\frac{Dt}{\pi}}. \tag{4.13}$$

This same result can also be derived, but more laboriously, in the lattice version of the SEP. As a general rule, it is almost always easier to solve random-walk-like processes in the continuum limit than in the discrete setting.

4.2 Asymmetric exclusion process

In the asymmetric exclusion process (ASEP), particle motion is driven by an overall spatial bias. We focus on the one-dimensional case since it is again the most interesting. Generally for the ASEP, a particle hops to the right with rate p and to the left with rate q; the hop is allowed only if the target site is empty. It is customary to set the total hopping rate to one, so that $p + q = 1$, and consider rightward bias: $p > q$. We shall tacitly assume that the hopping rates are fully asymmetric, $(p, q) = (1, 0)$, if not stated otherwise. The reason for this choice is that essentially all qualitative features of the ASEP are bias independent as long as some bias exists. Thus we do not lose in understanding but we typically gain in more compact formulae by considering the fully asymmetric model.

Steady states

The state of the ASEP at any moment is specified by the locations of all the particles. The probability $P(\mathcal{C}, t)$ to be in configuration $\mathcal{C}$ at time t evolves according to

$$\frac{dP(\mathcal{C}, t)}{dt} = \sum_{\mathcal{C}'} P(\mathcal{C}', t)\, R(\mathcal{C}' \to \mathcal{C}) - P(\mathcal{C}, t)\sum_{\mathcal{C}'} R(\mathcal{C} \to \mathcal{C}'). \tag{4.14}$$

Here $R(\mathcal{C}' \to \mathcal{C})\, dt$ is the transition probability from $\mathcal{C}'$ to $\mathcal{C}$ during a time interval dt. The system eventually reaches a steady state where

$$\sum_{\mathcal{C}'} P_{\mathrm{ss}}(\mathcal{C}')\, R(\mathcal{C}' \to \mathcal{C}) = P_{\mathrm{ss}}(\mathcal{C}) \sum_{\mathcal{C}'} R(\mathcal{C} \to \mathcal{C}')\,. \tag{4.15}$$

The above equations of motion are general and apply to any type of exclusion process on any lattice. There are, however, some subtle points that may cause confusion. For instance, shouldn't it be possible to use the basic postulate of equilibrium statistical physics to express the weight of the configuration $\mathcal{C}$ through its energy $E(\mathcal{C})$, i.e. to employ the equilibrium Gibbs distribution

$$P_{\mathrm{eq}}(\mathcal{C}) = Z^{-1}\, e^{-E(\mathcal{C})/T}, \tag{4.16}$$

where T is the temperature and $Z = \sum_{\mathcal{C}} e^{-E(\mathcal{C})/T}$ the partition function? The answer is generally *no*, for the simple reason that out-of-equilibrium systems usually don't have an energy function $E(\mathcal{C})$. Instead, such systems are usually defined by their dynamical rules. These rules lead to the transition rate matrix $R(\mathcal{C} \to \mathcal{C}')$ and to find the steady state we need to find a solution to the stationarity relation, Eq. (4.15). Steady states are therefore conceptually different from equilibrium states; to keep this property in mind we denote the underlying probabilities by $P_{\mathrm{ss}}(\mathcal{C})$ and $P_{\mathrm{eq}}(\mathcal{C})$, respectively.

The lack of an energy function is common in irreversible processes. For example, in aggregation, fragmentation, and adsorption, which we will study in the following chapters, the basic dynamical quantity is the number of particles, which is not conserved. This feature is a compelling sign that there is no energy function. In contrast, for conserved processes (like exclusion, where the particle number is conserved) it is sometimes possible to construct an energy function. In exclusion, the trivial energy function $E(\mathcal{C}) = 0$ if $\mathcal{C}$ is an allowed configuration and $E(\mathcal{C}) = \infty$ otherwise, does the job. The corresponding Gibbs distribution on the ring is simple: every allowed configuration has the same weight, as we shall show below. If so, why do we bother to derive the steady state on the ring from (4.15) rather than claim that it is a trivial consequence of the Gibbs formula? The reason is that the applicability of the Gibbs formula is accidental; almost any change in the microscopic rules of the exclusion process leads to steady states that are not given by the Gibbs formula.

Example 4.1. *Relaxed exclusion.* Consider ASEP on a ring and postulate that no more than two particles can be at the same site. Each particle hops with rate 1 to the neighboring site to the right, provided that this site is not overpopulated (i.e. it contains at most one particle); hopping to a site that contains two particles is forbidden. For instance, for a ring with $N = 3$ sites and $M = 3$ particles, the configuration $\{2, 1, 0\}$ is allowed and it evolves into $\{1, 2, 0\}$ (with rate 2 since either of the two particles can hop) or into $\{2, 0, 1\}$ (with rate 1). The governing equations for the configurational probabilities are:

$$\dot{P}(2, 1, 0) = P(1, 1, 1) - 3P(2, 1, 0),$$
$$\dot{P}(1, 2, 0) = P(0, 2, 1) + 2P(2, 1, 0) - 2P(1, 2, 0),$$
$$\dot{P}(1, 1, 1) = 2P(2, 0, 1) + 2P(0, 1, 2) + 2P(1, 2, 0) - 3P(1, 1, 1).$$

In the steady state, the left-hand sides vanish and rotational invariance gives $P_{ss}(2, 1, 0) = P_{ss}(0, 2, 1) = P_{ss}(1, 0, 2)$ and $P_{ss}(1, 2, 0) = P_{ss}(0, 1, 2) = P_{ss}(2, 0, 1)$. Solving then gives

$$P_{ss}(2, 1, 0) = P_{ss}(0, 2, 1) = P_{ss}(1, 0, 2) = \frac{2}{21} ,$$

$$P_{ss}(1, 2, 0) = P_{ss}(0, 1, 2) = P_{ss}(2, 0, 1) = \frac{1}{7} ,$$

$$P_{ss}(1, 1, 1) = \frac{2}{7} .$$

In contrast, there are six ways to arrange the particles so that the site occupancies are $(2, 1, 0)$ and one unique way to achieve the occupancy $(1, 1, 1)$. Thus

$$P_{eq}(2, 1, 0) = \frac{1}{7} \quad \text{plus all permutations,} \qquad P_{eq}(1, 1, 1) = \frac{1}{7} .$$

While the weight values are specific to the current example, the crucial feature is that the steady-state weights are different from the equilibrium weights. Thus, in general, we cannot use the Gibbs distribution to describe steady states in the ASEP.

Another subtlety is that instead of dealing with the complicated steady-state condition (4.15) it is tempting to impose the stronger *detailed balance condition*

$$P_{ss}(\mathcal{C}_2) \, R(\mathcal{C}_2 \to \mathcal{C}_1) = P_{ss}(\mathcal{C}_1) \, R(\mathcal{C}_1 \to \mathcal{C}_2) . \tag{4.17}$$

A solution that obeys detailed balance automatically satisfies (4.15) and therefore yields a steady state; however, detailed balance generally is too strong a constraint. In particular, Eq. (4.17) may have no solutions even when steady states exist (and satisfy (4.15)). As a trivial example, consider the ASEP with one particle (i.e. a biased random walk) on a ring of N sites. In the steady state, all configurations (specified by the position n of the particle) occur with equal weights. Hence $P_{ss}(n) = 1/N$ for all $n = 1, \ldots, N$. However, the detailed balance condition (4.17) yields

$$P_{ss}(n) \, R(n \to n + 1) = P_{ss}(n + 1) \, R(n + 1 \to n) ,$$
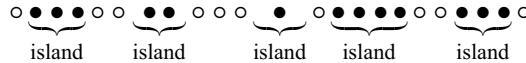
which is a contradiction, since $R(n \to n + 1) = 1$ and $R(n + 1 \to n) = 0$. Thus detailed balance fails for a fully biased random walk. This same argument shows that detailed balance is violated for any bias, $p \neq q$; detailed balance applies only for the symmetric random walk. A more concrete discussion of the distinction between the steady-state and detailed balance conditions is given in the highlight on page 186 of Section 6.3.

Later in this book we shall study non-equilibrium systems that have a well-defined equilibrium counterpart. In these examples, steady states will be the same as the equilibrium states given by Eq. (4.16), and the detailed balance condition is typically valid in such systems. We shall also meet out-of-equilibrium systems without an energy function, for which the detailed balance condition happens to apply. As we shall learn, detailed balance is a much stronger requirement than what is needed to ensure a steady state. While

equilibrium solutions and the detailed balance condition are not necessarily valid in out-of-equilibrium systems, when they do apply, they are potent and greatly help in understanding non-equilibrium systems.

Steady states on the ring

We now use the basic stationarity condition (4.15) to derive the weights of allowed configurations on a ring. Consider an allowed configuration $\mathcal{C}$ and denote by $I(\mathcal{C})$ the total number of islands in this configuration. For example, the configuration

$$\underbrace{\circ\,\bullet\,\bullet\,\bullet}_{\text{island}}\,\circ\,\circ\,\underbrace{\bullet\,\bullet}_{\text{island}}\,\circ\,\circ\,\circ\,\underbrace{\bullet}_{\text{island}}\,\circ\,\underbrace{\bullet\,\bullet\,\bullet\,\bullet\,\bullet}_{\text{island}}\,\circ\,\circ\,\underbrace{\bullet\,\bullet\,\bullet}_{\text{island}}\,\circ$$

contains $M = 13$ particles with $I = 5$ islands on a ring of size $N = 23$. In a time interval dt, the probability of leaving configuration $\mathcal{C}$ is equal to $I(\mathcal{C})\,P_{\text{ss}}(\mathcal{C})\,dt$. Similarly the probability of creating configuration $\mathcal{C}$ is $dt\sum_{\mathcal{C}'}P_{\text{ss}}(\mathcal{C}')$ where the sum is over configurations $\mathcal{C}'$ that differ from $\mathcal{C}$ by the position of one particle. That is, each $\mathcal{C}'$ is obtained from $\mathcal{C}$ by taking an island and moving its leftmost particle one step to the left. The total number of such configurations is therefore equal to the total number of islands $I(\mathcal{C})$. If all configurations have equal weights, $\sum_{\mathcal{C}'}P_{\text{ss}}(\mathcal{C}') = I(\mathcal{C})\,P_{\text{ss}}(\mathcal{C})$, and the stationarity condition is indeed obeyed. The total number of allowed configurations is $\binom{N}{M}$, which is just the number of ways of occupying N sites by M particles. Hence each allowed configuration has weight

$$P_{\text{ss}}(\mathcal{C}) = \binom{N}{M}^{-1}. \tag{4.18}$$

Let us now compute the current in the steady state. A configuration $\mathcal{C}$ contributes to the current through the bond between sites 1 and 2 if site 1 is occupied and site 2 is empty. Hence the following $N - 2$ sites $3, 4, \ldots, N$ must contain $M - 1$ particles. The total number of such configurations is $\binom{N-2}{M-1}$. Each of them contributes 1 to current (as the hopping rate is set to one), and the weight of each configuration is given by (4.18). Hence the steady-state current is

$$J = \frac{\binom{N-2}{M-1}}{\binom{N}{M}} = \frac{M}{N}\frac{N - M}{N - 1}. \tag{4.19}$$

Taking the thermodynamic limit $N, M \to \infty$ with $M/N = \rho$, we see that the steady-state current J and the density ρ are related by

$$J = \rho(1 - \rho). \tag{4.20}$$

The maximal current is $J_{\text{max}} = 1/4$ at $\rho_* = 1/2$. At higher densities, the current is less than J_{max} due to the small number of vacant sites (congested traffic); at lower densities, the current is also less than J_{max} due to the small density of particles (dilute traffic).

A remarkable feature of the macroscopic steady-state current (4.20) is that it agrees with the most naive expectation; namely, the current equals the product of the probabilities of

finding an occupied site (ρ) and a vacant site to the right ($1 - \rho$). The probability for such a configuration actually equals $\langle n_i(1 - n_{i+1}) \rangle$, where n_i is the occupation number at site i ($n_i = 0$ if i is empty and $n_i = 1$ if i is occupied). This expression gives the current j_i on bond $(i, i + 1)$. The factorization property,

$$j_i = \langle n_i(1 - n_{i+1}) \rangle = \langle n_i \rangle \langle (1 - n_{i+1}) \rangle = \rho(1 - \rho),$$

that holds for the ASEP, is a bit of a miracle. Essentially any alteration of the microscopic hopping or exclusion rules invalidates this factorization property.

4.3 Hydrodynamic approach

To understand the large-scale and long-time properties of exclusion processes, it is unnecessary to follow the motion of all the particles individually. This same feature underlies hydrodynamics, where we do not need to follow individual molecular trajectories and collisions to determine the motion of a bulk fluid. Thus instead of a microscopic description in terms of the positions and velocities of all the molecules, we use a macroscopic description that is based on the averaged "hydrodynamic" variables. An example of such a variable is the density $\rho(\mathbf{x}, t)$, which is defined as a total number of molecules in a *mesoscopic* domain that contains $\mathbf{x}$ divided by this domain volume. Here the term mesoscopic means:

- the domain should be sufficiently large to contain a large number of molecules so that relative fluctuations in this number are negligible;
- the domain should be sufficiently small so that the density (or any other hydrodynamic variable) is nearly constant inside the domain.

A hydrodynamic description is therefore sensible if the macroscopic scale, for example, the size of a body that is immersed in a gas flow, greatly exceeds the microscopic mean free path of the gas.

In addition to the density, the hydrodynamic description also involves the current $\mathbf{J}(\mathbf{x}, t)$ or the hydrodynamic velocity $\mathbf{v}(\mathbf{x}, t)$, and often a small number of other macroscopic variables, such as the pressure $p(\mathbf{x}, t)$ and the temperature $T(\mathbf{x}, t)$. These variables are related via hydrodynamic equations; the most basic is the mass conservation statement that has the form of the *continuity equation*

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{J} = 0, \qquad \mathbf{J} = \rho \mathbf{v}. \tag{4.21}$$

For the ASEP in one dimension, there are just two hydrodynamic quantities: the density $\rho(x, t)$ and the current $J(x, t)$. These are related by (4.20) and the one-dimensional version of the continuity equation (4.21). These equations constitute a complete *hydrodynamic* description, and represent the analog of the Euler formulation of classical hydrodynamics.

However, the hydrodynamics of the ASEP is much simpler than gas dynamics, as the former involves one quantity $\rho(x, t)$ that depends on two variables, x and t; the Euler

description is a system of coupled nonlinear partial differential equations for five[1] quantities: $\rho(\mathbf{x}, t)$ and $p(\mathbf{x}, t)$, plus three velocity components $\mathbf{v}(\mathbf{x}, t)$, that depend on four variables $\mathbf{x}, t$. In contrast, we shall see that the hydrodynamics of the ASEP is compact and tractable. The most important aspect of the macroscopic approach, however, is that it teaches us generic features of hydrodynamics. Indeed, a number of specific examples of the ASEP have been exactly solved at the microscopic level. These exact solutions allow us to compare with predictions of the hydrodynamic approach and learn about the validity of the latter. There is no such luxury in classical gases and fluids where exact microscopic solutions have never been found.

To see that the hydrodynamics of the ASEP is tractable, we combine (4.20) and the one-dimensional version of (4.21) to yield the first-order nonlinear partial differential equation (PDE)

$$\frac{\partial \rho}{\partial t} + (1 - 2\rho) \frac{\partial \rho}{\partial x} = 0. \tag{4.22}$$

The transformation $u = 1 - 2\rho$ recasts (4.22) into the most famous first-order nonlinear PDE, the *inviscid Burgers equation* $u_t + uu_x = 0$. In what follows, we shall also tacitly assume that the density lies in the range $0 \leq \rho \leq 1$, as implied by connection with the ASEP.

All first-order PDEs are hyperbolic (that is, wave-like), and propagate signals at finite speeds along *characteristics*. For Eq. (4.22), the characteristic is a curve in the *x–t* plane with slope

$$\frac{dx}{dt} = 1 - 2\rho. \tag{4.23}$$

Differentiating ρ along the characteristic, we find

$$\frac{d\rho}{dt} = \frac{\partial \rho}{\partial t} + \frac{dx}{dt} \frac{\partial \rho}{\partial x} = 0,$$

indicating that ρ does not change along the characteristic. (This is the reason why (4.23) was used to define the characteristic.) Since ρ is constant along the characteristics, the slope (4.23) is also constant, so the characteristics curves are straight lines.[2] The determination of the characteristics is the key to solving the general problem. Indeed, suppose that we want to solve the initial value problem, i.e. Eq. (4.22) subject to the initial condition $\rho(x, t = 0) = f(x)$. Let the (as yet unknown) characteristic begin at $x = \xi$ at time $t = 0$. Along this characteristic the density remains constant, $\rho = f(\xi)$, and therefore (4.23) becomes

$$\frac{dx}{dt} = 1 - 2f(\xi), \qquad x(t = 0) = \xi, \tag{4.24}$$

---

[1] For gases, we can express temperature via density and pressure and hence only two out of three thermodynamic quantities are independent.

[2] More generally, for homogeneous *quasi-linear* first-order PDEs, i.e. for equations of the form $\rho_t + c(\rho)\rho_x = 0$ with arbitrary $c = c(\rho)$, the characteristic has slope $dx/dt = c(\rho)$ and is a straight line.

whose solution is

$$x = \xi + [1 - 2f(\xi)]t. \tag{4.25}$$

The above formulae, in principle, complete the solution: on each characteristic curve (4.25) the density is $\rho = f(\xi)$. As long as characteristics do not intersect, this solution is well behaved. However, the most interesting behavior occurs when characteristics intersect. At an intersection, smooth or classical solutions cease to exist and a shock wave is born. The construction of such discontinuous shock-wave solutions is revealing as we shall illustrate below. Solutions of hyperbolic PDEs with shock waves are not at all pathological. In fact, the opposite is usually true, namely, solutions without shock waves are exceptional. For example, in the solution to Eq. (4.22), if there is a pair of points $\xi_1 < \xi_2$ such that the corresponding initial densities satisfy $f(\xi_1) < f(\xi_2)$, then the corresponding characteristics (4.25) intersect at time

$$t = \frac{1}{2} \frac{\xi_2 - \xi_1}{f(\xi_2) - f(\xi_1)}.$$

An initial density profile that is a monotonically decreasing function of the spatial variable is the only way to avoid the formation of a shock wave. Thus for the inviscid Burgers equation, at least, solutions to generic non-monotonic initial conditions do develop shocks.

Since discontinuous solutions are the rule rather than the exception, it is natural to study the evolution from discontinuous initial conditions. The classic example is:
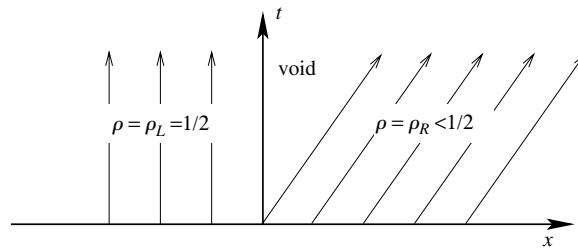
$$\rho_0(x) = \begin{cases} \rho_L, & x < 0, \\ \rho_R, & x > 0. \end{cases} \tag{4.26}$$

Let us determine how the system evolves starting from this initial step. Part of the reason for delving into this simple example is its many important applications. As one example, the evolution of the density step also solves the evolution of the Ising "corner interface" that will be discussed in Section 8.3.
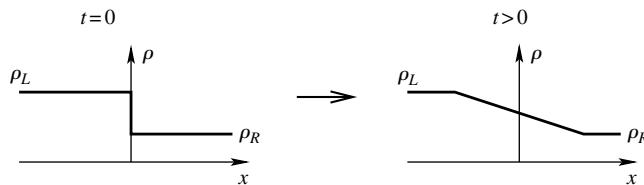
Example 4.2. *Rarefaction wave*: $\rho_L > \rho_R$. Since the initial density is a non-increasing function of x, there should be no shock waves. All the characteristics starting at $x < 0$ have the same slope $1 - 2\rho_L$; therefore $\rho = \rho_L$ in the region $x < (1 - 2\rho_L)t$. Similarly $\rho = \rho_R$ when $x > (1 - 2\rho_R)t$. We need to determine the density in the (infinite) triangular void region $(1-2\rho_L)t \leq x \leq (1-2\rho_R)t$ (Fig. 4.4). The solution should be continuous and it must connect $\rho = \rho_L$ to $\rho = \rho_R$. Taking $\rho = c$ with $\rho_L > c > \rho_R$ the corresponding characteristics (that must emanate from the origin) are defined, from Eq. (4.23), by $x = (1 - 2c)t$. Thus

$$\rho(x,t) = \begin{cases} \rho_L, & x < (1 - 2\rho_L)t, \\ \frac{1}{2}(1 - x/t), & (1 - 2\rho_L)t \leq x \leq (1 - 2\rho_R)t, \\ \rho_R, & x > (1 - 2\rho_R)t. \end{cases} \tag{4.27}$$

The solution of this type is called a rarefaction wave, or a fan (Fig. 4.5).

Fig. 4.4. Space-time diagram that illustrates a characteristic void. Shown is the case where $\rho_L = 1/2$ in which the initial data $\rho = \rho_L$ propagates along vertical characteristics, while the initial data $\rho = \rho_R$ propagates along characteristics of positive slope.


Fig. 4.5. Evolution of a density step into the rarefaction wave given by Eq. (4.27).

This same solution can also be obtained by a scaling approach. In the equation of motion (4.22), the same powers of x and t appear. This fact suggests that the solution should have the form $\rho(x,t) = f(z)$, with $z = x/t$. Substituting this form into Eq. (4.22) gives the ordinary differential equation $f'[(1 - 2f) - z] = 0$, and the solutions are either $f' = 0$ or $f = (1 - z)/2$. Ensuring continuity of these solutions, we reproduce Eq. (4.27).

Example 4.3. *Shock wave*: $\rho_L < \rho_R$. Here, the characteristics immediately intersect so that the initial shock wave persists. If c is the speed of the shock, the solution will maintain the initial step-function structure

$$\rho(x,t) = \begin{cases} \rho_L, & x < ct, \\ \rho_R, & x > ct. \end{cases}$$

To determine the shock speed c, we use mass conservation. Consider a large region that includes the shock. The current into this region from the left is $\rho_L(1 - \rho_L)$ and the current from this region is $\rho_R(1 - \rho_R)$. Equating the total mass flow into this region to the net increase of mass inside this region in a time Δt, we obtain

$$[\rho_L(1 - \rho_L) - \rho_R(1 - \rho_R)]\,\Delta t = (\rho_L - \rho_R)c\,\Delta t,$$

from which the speed of the shock wave is

$$c = 1 - \rho_L - \rho_R. \tag{4.28}$$

An intriguing aspect of the shock-wave solution arises when $0 < \rho_L < \rho_R$ and $\rho_R + \rho_L = 1$, for which the shock-wave speed $c = 1 - \rho_L - \rho_R = 0$. In this case, hydrodynamics does not fully predict the behavior near the shock. Since the underlying ASEP is a stochastic process, the location of the shock is also stochastic. In general, the shock is at $ct + \mathcal{O}(\sqrt{t})$, with fluctuations of the order of $\sqrt{t}$ that vary from realization to realization. Thus in the $t \to \infty$ limit, the shock wave is typically far from $x = 0$. If it happens to be far to the right, the density at $x = 0$ will equal ρ_L and this outcome happens in 50% of realizations. It is equally likely that the shock will be far to the left of $x = 0$ and the density will equal ρ_R. Thus for this special case, the density at the origin is a random variable, whereas it is deterministic for all other initial shock-wave conditions.

Let's now discuss the evolution of a piecewise constant initial density profile with two discontinuities:

$$\rho_0(x) = \begin{cases} \rho_L, & x < 0, \\ \rho_M, & 0 < x < X, \\ \rho_R, & x > X. \end{cases} \tag{4.29}$$

Now we have a scale X in the problem. The hydrodynamic approach is therefore applicable only when X is macroscopic, that is, it greatly exceeds the microscopic lattice spacing. Initially the discontinuities at $x = 0$ and X are far apart, so they do not "feel" each other. In the realm of hyperbolic PDEs, the above statement is literally correct because signals propagate with finite speeds. Thus the prediction for the evolution of step profile (4.26) provides the exact early-time behavior for the two-step initial condition – the result is a combination of solutions described in Examples 4.2 and 4.3. We now consider the behavior when the two excitations interact.

Example 4.4. *Shock-wave merging.* For $\rho_L < \rho_M < \rho_R$ the shock on the left has speed $1 - \rho_L - \rho_M$ that exceeds the speed $1 - \rho_M - \rho_R$ of the shock on the right. Therefore the left shock overtakes the right shock at time $T = X/(\rho_R - \rho_L)$. Subsequently, the shocks merge and form a single shock with speed $1 - \rho_L - \rho_R$. Thus for $t > T$ the solution becomes

$$\rho(x, t) = \begin{cases} \rho_L, & x < (1 - \rho_L - \rho_M)T + (1 - \rho_L - \rho_R)(t - T), \\ \rho_R, & x > (1 - \rho_L - \rho_M)T + (1 - \rho_L - \rho_R)(t - T). \end{cases}$$

Other possibilities include the interaction of two rarefaction waves, and the interaction of a shock and a rarefaction wave. We now consider an example of the latter.

Example 4.5. *Interaction of a shock and a rarefaction wave.* We assume that the shock is initially on the left, so that $\rho_L < \rho_M$, and that the second discontinuity generates the rarefaction wave, which requires $\rho_M > \rho_R$. For concreteness we set $\rho_L < \rho_R$. The shock wave initially moves with speed $c = 1 - \rho_L - \rho_M$, i.e. its position is $s = ct$. The rarefaction wave is non-trivial in the region $c_M t < x - X < c_R t$ where $c_M = 1 - 2\rho_M$ and $c_R = 1 - 2\rho_R$. Note that $c > c_M$ so that the shock wave collides with the rarefaction wave at time $T_* = X/(c - c_M) = X/(\rho_M - \rho_L)$. Then the interaction begins and it lasts until the

shock wave overtakes the entire rarefaction wave. During this stage, the position $s(t)$ of the shock satisfies

$$\frac{ds}{dt} = 1 - \rho_L - \frac{1}{2}\left(1 - \frac{s - X}{t}\right).$$

Solving the above equation subject to $s(T_*) = c\,T_*$ yields $s = X + (1 - 2\rho_L)t - 2(\rho_M - \rho_L)\sqrt{t\,T_*}$. At a later time T^* the shock reaches the right end of the rarefaction wave. From $s(T^*) = X + c_R T^*$ we find
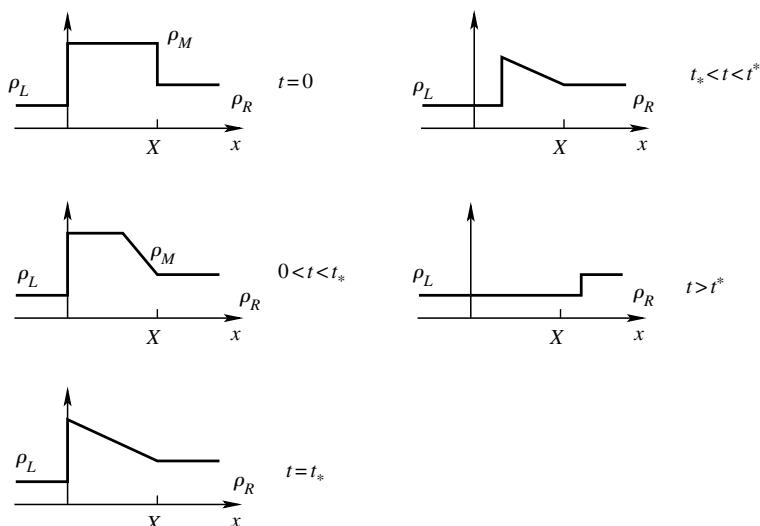
$$T^* = \left(\frac{\rho_M - \rho_L}{\rho_R - \rho_L}\right)^2 T_* = \frac{\rho_M - \rho_L}{(\rho_R - \rho_L)^2}\,X.$$

Finally, when $t \geq T^*$ the rarefaction wave has been swallowed by the shock wave and the outcome is a single shock that separates densities ρ_L and ρ_R. Thus the shock wave moves according to

$$s(t) = \begin{cases} (1 - \rho_L - \rho_M)t, & 0 \leq t \leq T_*, \\ X + (1 - 2\rho_L)t - 2\sqrt{X(\rho_M - \rho_L)t}, & T_* \leq t \leq T^*, \\ -\dfrac{\rho_M - \rho_R}{\rho_R - \rho_L}X + (1 - \rho_L - \rho_R)t, & t \geq T^*. \end{cases}$$

An illustration of the evolution of this system is shown in Fig. 4.6.

We can continue with the hydrodynamic approach and obtain a variety of useful results about density profiles. However, we now turn to our major goal of comparing exact results on the microscopic level with macroscopic solutions.


Fig. 4.6. Evolution of shock and rarefaction waves for the specific case of $\rho_L = \frac{1}{4}$, $\rho_M = \frac{3}{4}$, and $\rho_R = \frac{1}{2}$.

4.4 Microscopic approach

Because the ASEP is a strongly interacting system with infinitely many particles, it is not surprising that it has been exactly solved only in very few cases. However, the simple, single-step initial condition (4.26) is amenable to exact approaches.[3] The analysis is still involved and requires formidable tools, so here we present just a few key findings and qualitative results. From the hydrodynamic description of the two generic examples considered previously – the shock wave and the rarefaction wave – the shock wave is simpler to describe because it is essentially stationary, while the rarefaction wave continues to evolve. The same qualitative distinction also arises in a microscopic description, and the evolving rarefaction wave is much more difficult to analyze.

Shock wave

In the hydrodynamic approximation, a shock wave attains a (non-equilibrium) stationary state in a reference frame that moves with the shock, a feature that simplifies its analysis. For the ASEP, however, the location of the shock fluctuates from realization to realization. Thus we need to find a way to eliminate these fluctuations and keep the reference frame exactly at the shock location. This goal can be achieved by the clever trick of replacing one vacancy by a second-class particle and devising hopping rules that drive this second-class particle toward the location of the shock. The dynamical rules

$$10 \longrightarrow 01, \qquad 20 \longrightarrow 02, \qquad 12 \longrightarrow 21, \tag{4.30}$$

accomplish this goal. Here particles are denoted by 1, vacancies by 0, and the second-class particle is denoted by 2. All processes in (4.30) occur with rate 1, while processes that are not displayed, such as $21 \longrightarrow 12$, are forbidden. Each particle treats a second-class particle as a vacancy and hence the ASEP (the particle motion) is unaffected. Conversely, with respect to the vacancies, the second-class particle behaves as the regular particle.

Consider now the motion of a second-class particle with the step initial condition (4.26), and with $\rho_L < \rho_R$ to ensure a propagating shock wave. A second-class particle hops right with rate 1 if its right neighbor is a vacancy and hops left with rate 1 if its left neighbor is a particle. Thus if a second-class particle is in a uniform environment of particle density ρ, its speed equals the density of vacancies in front, $1 - \rho$, minus the density of particles behind, ρ; that is, $v = 1 - 2\rho$. For the step initial condition, the speed of a second-class particle is $1 - 2\rho_L$ to the left of the shock and $1 - 2\rho_R$ to the right of the shock. The former is larger than the shock speed $c = 1 - \rho_L - \rho_R$, while the latter is slower. Consequently, a second-class particle is attracted to the shock and we may define the position of the shock as the position of the second-class particle.

The above discussion makes plausible that the density profile in the reference frame of the second-class particle describes the shock microstructure. This was proved to be correct,

---

[3] However, no exact results are known for the double-step initial condition (4.29).

and the resulting density has been computed using the matrix ansatz by a technical and cumbersome calculation. However, some features of localized excitations can be determined by simpler reasoning. We discuss one such non-equilibrium steady state – the complementary rarefaction wave – in the next subsection.

Rarefaction wave

Consider, for concreteness, the initial condition $\rho_0(x) = 1$ for $x \leq 0$ and $\rho_0(x) = 0$ for $x > 0$, for which the hydrodynamic solution is

$$\rho(x,t) = \begin{cases} 1, & x < -t, \\ \frac{1}{2}\left(1 - x/t\right), & |x| \leq t, \\ 0, & x > t. \end{cases} \tag{4.31}$$

This hydrodynamic solution (4.31) has a jump in the first derivative at $x = \pm t$. These singularities are an outcome of the hydrodynamic approximation, namely the hyperbolic nature of the inviscid Burgers equation. However, the probability distribution $\mathcal{R}_M(t)$ that the rightmost particle is at site M at time t is a smooth function. We can determine $\mathcal{R}_M(t)$ by noting that the rightmost particle is unaffected by all other particles, so its location reduces to that of a biased random walk.[4] The probability distribution $\mathcal{R}_M(t)$ satisfies

$$\frac{d\mathcal{R}_M}{dt} = \mathcal{R}_{M-1} - \mathcal{R}_M,$$

whose solution, when the rightmost particle is initially at the origin, $\mathcal{R}_M(0) = \delta_{M,0}$, is

$$\mathcal{R}_M(t) = \frac{t^M}{M!}\, e^{-t}\,. \tag{4.32}$$

From this Poisson distribution we obtain

$$\langle M \rangle = \sum_{M \geq 0} M\mathcal{R}_M(t) = t, \qquad \langle M^2 \rangle - \langle M \rangle^2 = t. \tag{4.33}$$

The distribution of the rightmost particle is centered at $M = t$, which agrees with the hydrodynamic prediction, and the width of this distribution grows diffusively with time.

   The hydrodynamic description can be improved to resolve the discontinuities in the first derivative at $x = \pm t$. That such a resolution exists seems plausible if we think of classical hydrodynamics. The inviscid Burgers equation (4.22) is akin to the Euler description in classical hydrodynamics, which admits shock waves and other discontinuous solutions. The Navier–Stokes equations provide the next level of approximation in classical hydrodynamics, where such discontinuities are smoothed out. For the ASEP, the analog of the

---

[4] This argument relies on the assumption that the ASEP is totally asymmetric. In the partially asymmetric case the location of the left neighbor of the rightmost particle becomes important, which in turn is affected by its left neighbor, etc., so that the reduction to a single-particle problem is no longer valid.

Navier–Stokes equations is the Burgers equation

$$\frac{\partial \rho}{\partial t} + (1 - 2\rho)\frac{\partial \rho}{\partial x} = \frac{1}{2}\frac{\partial^2 \rho}{\partial x^2}, \tag{4.34}$$

whose solution for the rarefaction wave is indeed smooth. In the long-time limit this smooth solution is almost indistinguishable from (4.31) when x/t lies inside $(-1, 1)$, but it starts to deviate from the approximate hydrodynamic solution within a range $\pm\sqrt{t}$ about $x = \pm t$. The qualitative nature of this deviation is in accord with the exact prediction (4.33) for the width.

   As previously mentioned in the discussion of the shock wave, the classical methods to determine the density profile for the rarefaction wave are technical. Fortunately, some basic qualitative features of the rarefaction-wave solution can be grasped by heuristic reasoning. As an example, let us understand the behavior of the total current. We define the current as the total number of particles that have crossed the bond $(0, 1)$ up to time t, or equivalently the total number of particles $N(t)$ in the initially empty half-line $x > 0$. From the hydrodynamic solution (4.31), we find

$$N(t) = \int_0^\infty \rho(x, t)\, dx = \frac{1}{4}\, t\,. \tag{4.35}$$

Rather than integrating the density profile (4.31) we can understand (4.35) by using two basic features of the hydrodynamic solution: (i) the density at $x = 0$ is $1/2$ and (ii) the corresponding local current (see Eq. (4.20)) equals $J = \rho(1 - \rho) = 1/4$. Hence the time-integrated (or total) current is indeed $t/4$.

   Since $N(t)$ is a random quantity, Eq. (4.35) describes only the average total current. The challenge is to compute the current fluctuations. The primary feature is that current fluctuations scale as $t^{1/3}$, so that

$$N(t) = \frac{1}{4}\, t + t^{1/3}\xi \tag{4.36}$$

for large t, where ξ is a stationary random variable. To understand the $t^{1/3}$ scaling, we first map the ASEP onto a surface evolution model that is defined as follows (Fig. 4.7): only deposition events into local minima on the surface are allowed, and all such events
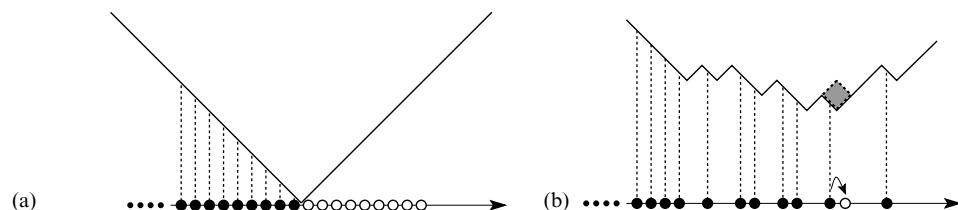

Fig. 4.7. Mapping of the ASEP onto a deposition model showing the single-step initial state (a) and the state at a later time (b). A deposition event (shaded square) corresponds to a hopping event in the ASEP (arrow).

occur at the same rate 1. The time-integrated current can be identified with the height of the interface at the origin. Equation (4.36) now admits a simple interpretation: the interface advances at speed 1/4 and fluctuations scale as $t^{1/3}$. The deposition model should be in the KPZ universality class (Section 2.9), and therefore the fluctuating part of Eq. (4.36) is merely a reformulation of KPZ scaling. As mentioned earlier, this mapping between the ASEP and surface evolution can be exploited to solve the corner interface problem in the Ising–Glauber model (Section 8.6).

To determine the distribution $\mathcal{P}(\xi)$ of the random variable ξ is a difficult problem, as $\mathcal{P}(\xi)$ happens to be a solution of the Painlevé II equation.[5] As a much easier alternative, we employ heuristic arguments to show that the extreme behaviors of the distribution $\mathcal{P}(\xi)$ are non-Gaussian. Let us first determine the asymptotic behavior of $\mathcal{P}(\xi)$ in the $\xi \to -\infty$ limit. Formally, we want to compute

$$\mathbb{P}(N, t) \equiv \text{Prob}[N(t) = N]$$

when $t/4 - N \gg t^{1/3}$. Consider the extreme case of no current, corresponding to $N = 0$. For this case, we must ensure that the rightmost particle has not moved; this criterion gives $\mathbb{P}(0, t) = e^{-t}$. Now the value of ξ corresponding to $N = 0$ is $\xi = -t^{2/3}/4$. Rewriting e^{-t} as $e^{-8|\xi|^{3/2}}$, we conclude that $\ln \mathcal{P}(\xi) \sim -c_-|\xi|^{3/2}$. As discussed in the highlight below, the exponent value in the asymptotic behavior of $\mathcal{P}(\xi)$ that is obtained by this heuristic approach is generally correct, but the amplitude that multiplies the factor $|\xi|^{3/2}$ is unreliable. Thus we write the asymptotic behavior as $\ln \mathcal{P}(\xi) \sim -c_-|\xi|^{3/2}$, with unknown amplitude c_- rather than the factor 8 that is given by the heuristic approach.

Probability distributions from tail behavior

The estimates given in Eqs (4.42) are illustrations of a general and handy method to determine the probability distribution of a stochastic quantity by focusing on its extremal behavior. These extremal properties are generally easy to determine, from which the leading behavior of the probability distribution can be inferred with minimal effort.

As a first and trivial example, consider the probability distribution of a random walk in one dimension. Suppose that we don't know that this distribution is Gaussian, but that we do know that the typical displacement grows with time as $t^{1/2}$. Given this fact, then under the assumption of scaling, the probability distribution $P(x, t)$ should have the form

$$P(x, t) \simeq \frac{1}{\sqrt{t}} \Phi(x/\sqrt{t}), \tag{4.37}$$

where the prefactor imposes normalization, $\int P(x, t)\, dx = 1$. What is the scaling function $\Phi(z)$? This function decays for large z, but how fast? Clearly, $\Phi(z)$ must decay faster than any power law for large z. If not, sufficiently high moments of the distribution would diverge.

---

[5] The Painlevé II equation is $y''(x) = 2y^3 + xy + \alpha$, where α is a complex constant.

A natural hypothesis then is

$$\Phi(z) \sim \exp(-z^\delta)\,. \tag{4.38}$$

Using this form, we determine the "tail" exponent δ as follows: Consider a "stretched" walk in which every step is to the right. The probability of making t consecutive such steps is 2^{-t}. On the other hand, this stretched walk corresponds to the scaling variable $z = x/t^{1/2} = t^{1/2}$, and Eq. (4.38) gives $\Phi(z = t^{1/2}) \sim \exp[-(t^{1/2})^\delta]$. To match with the exact behavior 2^{-t}, we must have $\delta = 2$. Thus we conclude that $\Phi(z) \sim \exp(-\text{const.} \times z^2)$. Because of the crudeness of the estimate for the tail probability, we do not know the constant inside this Gaussian function.

As a slightly less trivial example, consider a self-avoiding walk in d dimensions; this is a lattice random walk with the additional constraint any site can be visited no more than once. The typical displacement of the self-avoiding walk grows as t^ν; it is known that $\mu = 1/2$ for $d > 4$, so that self-avoidance is asymptotically negligible, while $\nu > 1/2$ for $d < 4$. In analogy with Eq. (4.37), we expect that the distribution of endpoints of the walk at time t is given by

$$P(r, t) \to t^{-d\nu}\,\exp\left[-a(r/t^\nu)^\delta\right]\,. \tag{4.39}$$

Consider again a stretched self-avoiding walk in which all t steps are in the same direction. This stretched walk corresponds to the scaling variable $z = x/t^\nu = t^{1-\nu}$ so that the endpoint distribution becomes

$$P(r \sim t, t) \to t^{-d\nu}\,\exp\left[-a(t^{1-\nu})^\delta\right]\,. \tag{4.40}$$

On the other hand, this stretched walk occurs with probability $(2d)^{-t}$ since all steps must be in the same direction. By comparing $(2d)^{-t}$ with (4.40), we obtain the general exponent relation

$$\delta = (1 - \nu)^{-1}\,. \tag{4.41}$$

We used this same reasoning to give the estimates in Eq. (4.42). This approach works in a wide variety of contexts and has proven to be powerful in its consequences, yet simple to apply.

For the large-ξ behavior, we need to estimate $\mathbb{P}(N, t)$ in the region $N - t/4 \gg t^{1/3}$. For this purpose, we modify the ASEP in an inconsequential manner so that N has a finite upper limit. Instead of a continuous-time process, we define a discrete-time evolution in which all "eligible" particles (those with vacant right-neighbor sites) hop with probability 1/2. For this synchronous ASEP, the total number of particles that move to the half-space $x > 0$ satisfies the bound $N \le t$. Let us now compute the probability of reaching this upper limit. On the first step, the single eligible particle must hop (with probability 1/2); on the second

step, there are two eligible particles that must hop (with probability 1/4), etc. Thus

$$\mathbb{P}(t,t) = \frac{1}{2} \times \frac{1}{2^2} \times \cdots \times \frac{1}{2^{t-1}} = 2^{-t(t-1)/2}.$$

The value of ξ corresponding to $N = t$ is $\xi = 3t^{2/3}/4$. Rewriting $2^{-t(t-1)/2}$ in terms of ξ we establish the asymptotic, $\ln \mathcal{P} \sim -\xi^3$. Again, the detailed properties of this asymptotic should not be trusted – both since ξ diverges simultaneously with t and because we modified the ASEP rules – but qualitatively the dependence on ξ should be correct.

Combining our results, the tails of the probability distribution are given by

$$\mathcal{P}(\xi) \sim \begin{cases} e^{-c_-|\xi|^{3/2}}, & \xi \to -\infty, \\ e^{-c_+\xi^3}, & \xi \to \infty. \end{cases} \tag{4.42}$$

Thus the distribution is asymmetric and the extreme behaviors are non-Gaussian.

4.5 Open systems

Finally we present an exact solution for a non-equilibrium steady state of an open system (Fig. 4.8) that connects two infinite particle reservoirs at different densities. The particles follow the ASEP rules in the bulk, while the leftmost site becomes occupied at rate α and the hopping rate out of the rightmost site is β. Symbolically,

$$0 \xrightarrow{\alpha} 1 \qquad \text{at site } i = 1, \qquad \qquad 1 \xrightarrow{\beta} 0 \qquad \text{at site } i = N.$$

Our goal is to determine the properties of the steady state.

Let n_i be the occupation number: $n_i = 1$ if site i is occupied and $n_i = 0$ if this site is empty. We shall also use the flux variables $j_i = n_i(1 - n_{i+1})$. The state of site i changes when either: (i) the site is vacant and its left neighbor is occupied, then the $0 \to 1$ transition occurs with rate j_{i-1}; or (ii) the site is occupied and its right neighbor is vacant, then the $1 \to 0$ transition occurs with rate j_i. Taking into account that $j_i = 0$ if $n_i = 0$ and $j_{i-1} = 0$ if $n_i = 1$ we see that

$$n_i \to 1 - n_i \qquad \text{with rate} \quad w_i = j_{i-1} + j_i. \tag{4.43}$$

Equivalently, in an infinitesimal time interval dt the occupation number evolves according to

$$n_i(t + dt) = \begin{cases} n_i(t), & \text{with probability } 1 - w_i \, dt, \\ 1 - n_i(t), & \text{with probability } w_i \, dt. \end{cases} \tag{4.44}$$
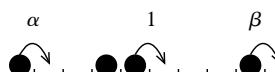
Fig. 4.8.   ASEP in an open system. One of the two bulk particles is (temporarily) immobile due to exclusion.

Averaging (4.44), the local density $\langle n_i \rangle$ evolves according to

$$\frac{d\langle n_i \rangle}{dt} = \langle w_i(1 - 2n_i) \rangle. \tag{4.45}$$

By using $w_i = j_{i-1} + j_i$ and the expressions for the local currents we recast (4.45) into

$$\frac{d\langle n_i \rangle}{dt} = \langle n_{i-1}(1 - n_i) \rangle - \langle n_i(1 - n_{i+1}) \rangle. \tag{4.46}$$

Intuitively, this equation expresses the change of the local density at a given site as the current into this site minus the current out of this site. However, single-site averages are coupled to two-site averages. If we write equations for local currents (two-site average), we obtain three-site averages. Continuing, one obtains an infinite equation hierarchy in which k-site averages are expressed in terms of $(k + 1)$-site averages. This hierarchical nature is the main obstacle that has forestalled an analytical solution for the time-dependent behavior.

The dynamical rule (4.43) also holds at the boundary if we set $j_0 = \alpha(1 - n_1)$ and $j_N = \beta n_N$. Using these results, together with the general equation

$$\frac{d}{dt}\langle n_i \rangle = \langle w_i[-n_i + (1 - n_i)] \rangle,$$

gives the evolution equations at the boundary sites. In conjunction with (4.46) we obtain the equations of motion

$$\frac{d\langle n_i \rangle}{dt} = \begin{cases} \alpha\langle 1 - n_1 \rangle - \langle n_1(1 - n_2) \rangle, & i = 1, \\ \langle n_{i-1}(1 - n_i) \rangle - \langle n_i(1 - n_{i+1}) \rangle, & 1 < i < N, \\ -\beta\langle n_N \rangle + \langle n_{N-1}(1 - n_N) \rangle, & i = N. \end{cases} \tag{4.47}$$

Matrix method

The steady state of the ASEP in the open system is exactly soluble using a matrix-product ansatz.[6] The main idea is to write the steady-state probability distribution in an abstract way as a product of matrices

$$P(\mathbf{n}) = \frac{1}{Z_N} \left\langle W \left| \prod_{i=1}^{N} [n_i D + (1 - n_i)E] \right| V \right\rangle. \tag{4.48}$$

Here $\mathbf{n} = (\ldots, n_{i-1}, n_i, n_{i+1}, \ldots)$ specifies the occupancy at each lattice site, D, E are matrices whose properties are given below, and $\langle W|, |V \rangle$ are state vectors. Inside the product, the matrix D is associated with an occupied site ($n_i = 1$) and the matrix E with an empty site. For example, we may write the probability for the occupancy configuration ($\{1, 0, 1, 0\}$ as $P(\{1, 0, 1, 0\}) = Z_4^{-1}\langle W|DEDE|V \rangle$. The factor Z_N in (4.48) ensures that

---

[6] This method is related to the Bethe ansatz, used to obtain exact solutions in equilibrium statistical mechanics.

the probability distribution is normalized; thus $\sum_{\mathbf{n}} P(\mathbf{n}) = 1$ after summing over all 2^N possible configurations of the N-site lattice. The normalization factor therefore equals

$$Z_N = \langle W | (D + E)^N | V \rangle \equiv \langle W | C^N | V \rangle, \tag{4.49}$$

with $C = D + E$. Using the matrix-product ansatz (4.48) we can express the local density and current in the compact forms

$$\begin{aligned} \langle n_i \rangle &= \langle W | C^{i-1} D C^{N-i} | V \rangle / Z_N, \\ \langle j_i \rangle &= \langle W | C^{i-1} D E C^{N-i-1} | V \rangle / Z_N. \end{aligned} \tag{4.50}$$

Thus for $\langle n_i \rangle$, for example, the first $i - 1$ and the last $N - i - 1$ sites must be empty, as encoded by the factors C^{i-1} and C^{N-i-1}, while site i is occupied, as mandated by the factor D. To implement the matrix method, we should find matrices D, E and vectors $\langle W |, | V \rangle$ such that the matrix-product ansatz (4.48) gives the steady state. In fact, we shall see that we do not need explicit representations of these matrices; the important feature is the algebra obeyed by these matrices.

In the steady state, Eqs (4.47) imply that the average local current is homogeneous: $\langle j_0 \rangle = \langle j_1 \rangle = \cdots = \langle j_i \rangle = \cdots = \langle j_{N-1} \rangle = \langle j_N \rangle$. These conditions yield

$$\alpha \langle W | E C^{N-1} | V \rangle = \langle W | D E C^{N-2} | V \rangle, \tag{4.51a}$$

$$\langle W | C^{i-1} D E C^{N-i-1} | V \rangle = \langle W | C^i D E C^{N-i-2} | V \rangle, \tag{4.51b}$$

$$\langle W | C^{N-2} D E | V \rangle = \beta \langle W | C^{N-1} D | V \rangle. \tag{4.51c}$$

The bulk condition (4.51b) is satisfied when $(DE)C = C(DE)$, i.e. the matrix DE commutes with $C = D + E$. This commutativity is assured if the matrices DE and $D + E$ are proportional to each other. It turns out that a solution in the matrix-product form (4.48) is applicable if we choose the simplest relation between DE and $D + E$, namely, $DE = D + E$. Substituting this result into (4.51a) and (4.51c) we find that these equations are indeed satisfied if $\alpha \langle W | E = \langle W |$ and $\beta D | V \rangle = | V \rangle$, respectively. Therefore if the matrices D, E and and the vectors $\langle W |$ and $| V \rangle$ obey

$$DE = D + E, \qquad \langle W | E = \alpha^{-1} \langle W |, \qquad D | V \rangle = \beta^{-1} | V \rangle, \tag{4.52}$$

then the local densities are indeed stationary.

The stationarity of the local densities does not prove that the matrix ansatz gives the steady state: we haven't shown that the local currents (or generally the two-point correlation functions $\langle n_i n_j \rangle$) are stationary, or that the three-point correlation functions are stationary, etc. So the matrix-product ansatz (4.48), together with relations (4.52), represent an educated guess that happens to be correct for the ASEP in the open-system setting. In particular, once the one-point averages are stationary, all higher-order averages and the entire distribution $P(\mathbf{n})$ are stationary under the dynamics.

The proof that the matrix ansatz gives the stationary distribution is straightforward, but tedious. We illustrate the validity of the matrix ansatz with two examples.

Example 4.6. *Small systems.* Consider the simplest case $N = 1$. Using (4.52) we get $\langle W|E|V \rangle = \alpha^{-1}\langle W|V \rangle$ and $\langle W|D|V \rangle = \beta^{-1}\langle W|V \rangle$. Thus the system is empty with probability $P(0) = \alpha^{-1}/(\alpha^{-1} + \beta^{-1}) = \beta/(\alpha + \beta)$ and occupied with probability $P(1) = \alpha/(\alpha + \beta)$. These probabilities are stationary. Since the empty state is destroyed with rate $\alpha P(0)$ and created with rate $\beta P(1)$, the equality $\alpha P(0) = \beta P(1)$ holds.

When $N = 2$, we use (4.52) to compute

$$\langle W|EE|V \rangle = \alpha^{-2}\langle W|V \rangle,$$

$$\langle W|ED|V \rangle = \alpha^{-1}\beta^{-1}\langle W|V \rangle,$$

$$\langle W|DE|V \rangle = (\alpha^{-1} + \beta^{-1})\langle W|V \rangle,$$

$$\langle W|DD|V \rangle = \beta^{-2}\langle W|V \rangle.$$

To verify, for example, that the empty state $\{0, 0\}$ is destroyed and created with equal rates, we must verify

$$\alpha\langle W|EE|V \rangle = \beta\langle W|ED|V \rangle.$$

This relation is indeed satisfied. Similarly, the three remaining states are stationary.

Example 4.7. *Special configurations in an arbitrary length system.* Consider the system of length N. Let us prove that the special configuration

$$\underbrace{0\ldots00}_{M}\underbrace{11\ldots1}_{N-M}, \tag{4.53}$$

whose probability is equal to $\langle W|E^M D^{N-M}|V \rangle/Z_N$ according to the matrix ansatz, is created and destroyed with equal rates. (Both N and M are arbitrary, modulo the constraint $1 \leq M \leq N - 1$.)

The configuration (4.53) is created at rate $\langle W|E^{M-1}DED^{N-M-1}|V \rangle/Z_N$, as it comes from the configuration

$$\underbrace{0\ldots0}_{M-1}10\underbrace{1\ldots1}_{N-M-1} .$$

Destruction of the state (4.53) occurs when a particle enters from the reservoir on the left or if the rightmost particle leaves the system. The total destruction rate is therefore $(\alpha + \beta)\langle W|E^M D^{N-M}|V \rangle/Z_N$. Thus the weight of configuration (4.53) remains constant if

$$\langle W|E^{M-1}DED^{N-M-1}|V \rangle = (\alpha + \beta)\langle W|E^M D^{N-M}|V \rangle. \tag{4.54}$$

Using (4.52) it is straightforward to verify (4.54).

An extension of the analysis of the previous example gives a proof that the solution given by the matrix ansatz is stationary. We are not yet done as we haven't proven the existence of the matrices D, E and the vectors $\langle W|, |V \rangle$ that satisfy (4.52). The best way to deal with

this problem is to ignore it: we simply postulate relations (4.52) and then all the weights are computed using pure algebraic manipulations. This approach is very much in the spirit of the raising and lowering operators for the harmonic oscillator in quantum mechanics, where we almost never need explicit representations for these operators – all that is needed is the operator algebra.

For explicit computations we meet matrix products such as $\langle W | \dots D^a E^b D^c E^d \dots | V \rangle$. Using the relation $DE = E + D$ we can move operators E to the left and rewrite matrix products as sums of canonical matrix products $\langle W | E^A D^B | V \rangle$. We can calculate the canonical matrix elements by using the remaining relations (4.52) to give

$$\langle W | E^p D^q | V \rangle = \alpha^{-p} \beta^{-q} \langle W | V \rangle. \tag{4.55}$$

Example 4.8. *Commuting matrices*. Let us explore whether it possible that matrices D and E commute, that is, $ED - DE = 0$. Using (4.52) we find the general expression for the matrix element of the commutator

$$\langle W | (DE - ED) | V \rangle = \left(\frac{1}{\alpha} + \frac{1}{\beta} - \frac{1}{\alpha\beta} \right) \langle W | V \rangle = \frac{\alpha + \beta - 1}{\alpha\beta} \langle W | V \rangle.$$

Hence matrices D and E can commute only when $\alpha + \beta = 1$. Along this line we can choose the matrices to be scalar, $E = \alpha^{-1}$ and $D = \beta^{-1}$, and the vectors to be $\langle W | = | V \rangle = 1$. Thus along the line $\alpha + \beta = 1$ there are no correlations; in particular, the configuration with M occupied sites occurs with probability $\alpha^M \beta^{N-M}$, independent of the locations of the particles. For these values of α and β, the open system has the same behavior as the closed ring.

Apart from the line $\alpha + \beta = 1$ along which the one-dimensional representation suffices, the representation must be infinite-dimensional! Here is a rare case where the solution of a finite problem (with 2^N states) requires infinite-dimensional matrices. We do not prove the above assertion and we do not give any explicit representation since, as stated already, there is no need for the explicit representation of matrices and vectors – the computations are easier to perform formally, merely using (4.52) as a set of axioms.

The steady state

Let us now compute the current $J = \langle j_i \rangle$. Using (4.50) for $\langle j_i \rangle$, in conjunction with $DE = D + E = C$, gives the compact general formula

$$J = \frac{\langle W | C^{N-1} | V \rangle}{\langle W | C^N | V \rangle}, \tag{4.56}$$

which tells us that we only need to compute $Z_N = \langle W | C^N | V \rangle$ for all N. For this task, we follow the strategy of representing the operator $(D + E)^N$ as a sum of operators in canonical form (with Es preceding Ds). For $N = 2$,

$$(D + E)^2 = E^2 + ED + DE + D^2$$
$$= E^2 + ED + (D + E) + D^2,$$

since only the term DE needed to be re-ordered. Similarly for $N = 3$,

$$
\begin{aligned}
(D + E)^3 &= E^3 + E^2D + EDE + DE^2 + ED^2 + DED + D^2E + D^3 \\
&= E^3 + E^2D + E(D + E) + (D + E)E + ED^2 \\
&\quad + (D + E)D + D(D + E) + D^3 \\
&= E^3 + E^2D + ED + E^2 + (D + E) + E^2 + ED^2 \\
&\quad + D^2 + ED + D^2 + (D + E) + D^3 \\
&= E^3 + E^2D + ED^2 + 2(E^2 + ED + D^2) + 2(E + D)
\end{aligned}
$$

is the required sum. Note that canonical terms with the same total degree have the same amplitude. This fact suggests the general pattern

$$
(D + E)^N = \sum_{p=1}^{N} A(p, N) \sum_{q=0}^{p} E^q D^{p-q} .
$$

It is easy to compute the amplitudes $A(p, N)$ for $p \leq N$ and small N, and for arbitrary N along the diagonal $p = N$, where $A(N, N) = 1$. To proceed further and determine the general behavior requires dedication and creativity. However, the result is known:

$$
(D + E)^N = \sum_{p=1}^{N} p \, \frac{(2N - 1 - p)!}{N!(N - p)!} \sum_{q=0}^{p} E^q D^{p-q} . \tag{4.57}
$$
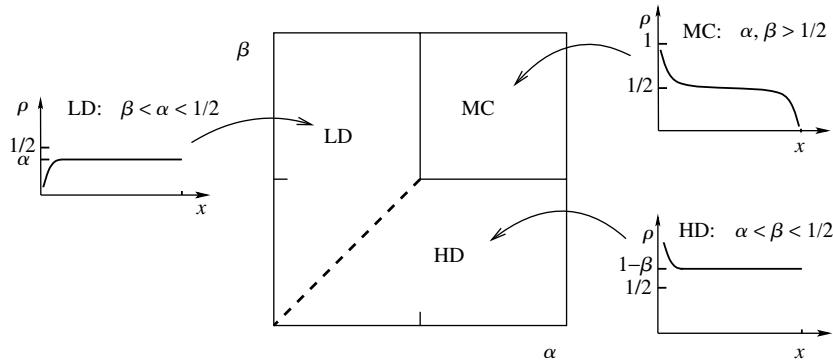
While the derivation of this result is anything but straightforward, its verification is straightforward. For example, one can prove (4.57) by induction.

The canonical matrix element is given by Eq. (4.55), which together with (4.57) leads to

$$
\frac{\langle W | C^N | V \rangle}{\langle W | V \rangle} = \sum_{p=1}^{N} p \, \frac{(2N - 1 - p)!}{N!(N - p)!} \frac{\beta^{-p-1} - \alpha^{-p-1}}{\beta^{-1} - \alpha^{-1}} . \tag{4.58}
$$

Equations (4.56) and (4.58) yield the current for the system of arbitrary length N, and with arbitrary incoming and outgoing rates α and β. While the problem is formally solved, the general solution is not very illuminating – the answer is the sum of ratios. In statistical physics, simple and striking behaviors emerge in the thermodynamic limit that are much more revealing than the behaviors of few-particle systems. Hence we probe the thermodynamic limit $N \to \infty$. In the present case, the emergent behavior is indeed simple and beautiful – there are hidden phases and phase transitions between these phases, as manifested by the following simple formula for the current in Eq. (4.56):

$$
J = \begin{cases} 1/4, & \alpha \geq 1/2, \ \beta \geq 1/2 \quad \text{(MC phase)}, \\ \alpha(1 - \alpha), & \alpha < 1/2, \ \beta > \alpha \quad \text{(LD phase)}, \\ \beta(1 - \beta), & \beta < 1/2, \ \alpha > \beta \quad \text{(HD phase)}. \end{cases} \tag{4.59}
$$

Fig. 4.9. Phase diagram of the ASEP. Also shown are schematic plots of the density versus position for the maximal-current (MC) phase, the high-density (HD) phase, and the low-density (LD) phase.

Here the abbreviations MC, LD, and HD refer to the maximal-current, low-density, and high-density phases, respectively. Recalling relation (4.20) between the current and the density, we see that the density in the MC phase is $\rho = 1/2$. A similar argument for the LD phase gives two solutions, $\rho = \alpha$ and $\rho = 1 - \alpha$; the exact solution shows that $\rho = \alpha$, so that the density is indeed low in the LD phase (since $\alpha < 1/2$). In the HD phase $\rho = 1 - \beta$. The lines $\alpha = 1/2, \beta > 1/2$ and $\beta = 1/2, \alpha > 1/2$ demarcate continuous transitions between adjoining phases (Fig. 4.9). Finally, the line $\alpha = \beta < 1/2$ represents a discontinuous transition between the LD and HD phases. In contrast to the MC, LD, and HD phases, which are characterized by a uniform bulk density, the density is a linear function of $x = i/N$ when $\alpha = \beta < 1/2$. In summary,

$$\rho = \begin{cases} 1/2, & \alpha \geq 1/2, \ \beta \geq 1/2 & \text{(MC phase)}, \\ \alpha, & \alpha < 1/2, \ \beta > \alpha & \text{(LD phase)}, \\ 1 - \beta, & \beta < 1/2, \ \alpha > \beta & \text{(HD phase)}, \\ \alpha + (1 - 2\alpha)x, & \alpha = \beta < 1/2. \end{cases} \tag{4.60}$$

Near the entrance and the exit, the local density departs from the above bulk values. Consider first the maximal-current phase. The deviation from the bulk is the largest when $\alpha = \infty$. In this case the leftmost site is always occupied, $n_1 = 1$, and hence the density decays from 1 to 1/2 as one moves away from the left boundary. The approach to the bulk density is algebraic, $\langle n_j \rangle - 1/2 \sim j^{-1/2}$ when $1 \ll j$. Similarly, near the output boundary the local density also decays algebraically, and the asymptotic behavior is

$$\frac{1}{2} - \langle n_{N-j} \rangle = \frac{1}{\sqrt{4\pi j}} + \mathcal{O}(j^{-3/2}) \tag{4.61}$$

when $1 \ll j \ll N$. Thus the leading asymptotic is independent[7] of the rates α and β. The derivation of (4.61) is straightforward: we have the exact solution and we just need to

---

[7] The same universality occurs on the left boundary.

perform an asymptotic analysis to estimate sums, such as the one appearing in (4.58). (Some hints for doing such calculations are mentioned in the problems.) In the maximal-current phase, mean-field theory correctly predicts bulk quantities, but gives a j^{-1} tail for the local density instead of the correct $j^{-1/2}$ tail in Eq. (4.61). The same picture arises in the other two phases, namely, mean-field theory gives the correct bulk behavior, but gives the incorrect behavior of the density near the boundaries.

Finally, let us try to understand the first-order phase transition between the LD and HD phases. Consider the limiting case $\alpha N \ll 1$ and $\beta N \ll 1$; also let $\alpha > \beta$. In the long-time limit, the system will mostly be in configurations such as $0011 \ldots 11$, with just a few empty sites on the left. To see why these configurations predominate, imagine that the rightmost particle has just exited the system. Then the entire occupied domain will move to the right; this event will take a time of the order of N, which is much smaller than the time $1/\alpha$ for a new particle to enter, or the time $1/\beta$ for a particle to leave. Similarly, if a new particle enters the system, the whole occupied domain moves one step to the left. Hence the domain wall between empty and occupied sites undergoes a biased random walk with negative drift velocity $\beta - \alpha$. Therefore the domain wall is biased toward the left boundary and the system is indeed in the high-density state. Similarly when $\alpha < \beta$, the system is typically in a configuration like $00 \ldots 0011$, with just a few occupied sites on the right. When $\alpha = \beta$, there is no drift; hence, the domain wall undergoes an unbiased random walk on an interval (with reflecting boundaries). Averaging, we conclude that the density profile is linear: $\langle n_i \rangle = i/N$.

4.6 Notes

The exclusion process was originally proposed to describe the diffusion of large molecules through narrow pores [65]. Since then, the model has been rediscovered by probabilists, including Harris [66] and Spitzer [67]. The method used in Section 4.1 to solve for the probability distribution of a tagged particle in the symmetric exclusion process is due to Levitt [68].

For the hydrodynamic approach to exclusion processes, the books by Burgers [56] and Whitham [69] are quite instructive. The necessary elements of the theory of non-linear hyperbolic partial differential equations (e.g. the nonlinear wave equation) are described in [56, 69] and in many other sources; a useful general reference is Logan's text [70].

Exclusion processes have been quite popular over the past 20 years. The book by Liggett [71] and the review articles by Schmittmann and Zia [72], Derrida and Evans [73], Schütz [74], Golinelli and Mallick [75], Blythe and Evans [76], and Derrida [77, 78] provide a wealth of information about the physical and transport properties of exclusion processes. For more details of the extremal argument to estimate the tail of a probability distribution from the highlight on page 120, see e.g. de Gennes' book on polymer physics [11]; this approach has also been coined the Lifshitz tail; see [79] for Lifshitz's contributions to this subject.

4.7 Problems

4.1 Consider the extreme limit where the lattice is fully occupied, apart from one vacancy that is initially at the origin. The problem is to investigate the displacement of the tracer particle.

(a) Show that in one dimension if the tracer particle is initially to the right of the vacancy, say at site $n > 0$, then at any time the tracer particle will be either at its original position n or at site $n - 1$. Estimate the typical number of hops of the tracer particle during the time interval $(0, t)$. Show that the probability that the tracer particle has not moved decays as $t^{-1/2}$ in the long-time limit.

(b) Try to answer the same questions as in (a) in the two-dimensional case. Argue that the typical number of hops scales as $\ln t$ and that the typical displacement grows[8] as $\sqrt{\ln t}$.

(c) Consider the same problem in three dimensions. Show that the tracer particle will never move with a positive probability. Argue that the number of hops throughout the entire evolution is finite.

4.2 Consider the inviscid Burgers equation (4.22) with the initial condition $\rho(x, 0) = (1 + x^2)^{-1}$. Show that a smooth solution exists up to time $t_* = 4/\sqrt{27}$, and that a shock wave originates at $x_* = -5/\sqrt{27}$ at time t_*.

4.3 Consider the interaction of a shock and a rarefaction wave that are generated starting with initial condition (4.29). Assume that $\rho_1 < \rho_2$ and $\rho_2 > \rho_3$. In contrast to Example 4.5, consider the case when $\rho_1 > \rho_3$.

4.4 Describe the evolution that arises from the initial condition (4.29) when $\rho_1 > \rho_2 > \rho_3$, in which two rarefaction waves interact.

4.5 Give a heuristic derivation of the Burgers equation (4.34) by the following steps:

(a) Start with the master equation

$$\frac{\partial \rho(x, t)}{\partial t} = \rho(x - 1, t)[1 - \rho(x, t)] - \rho(x, t)[1 - \rho(x + 1, t)].$$

Try to justify its validity using the exact evolution equation (4.46). What are the caveats in writing local currents as products?

(b) Expand the densities in Taylor series

$$\rho(x \pm 1, t) = \rho(x, t) \pm \frac{\partial \rho(x, t)}{\partial x} + \frac{1}{2} \frac{\partial^2 \rho(x, t)}{\partial x^2} + \cdots.$$

Substitute these expansions into the master equation of part (a) and deduce the Burgers equation (4.34).

---

[8] The probability distribution of the tracer particle in two dimensions is *not* Gaussian; however, the central-limit-type argument gives the correct $\sqrt{\ln t}$ scaling for the typical displacement.

4.6 Consider the Burgers equation in the canonical form $u_t + uu_x = Du_{xx}$. The goal of this problem is to solve the Burgers equation for an arbitrary initial condition.

(a) Use the Cole–Hopf transformation

$$u = -2D(\ln \phi)_x = -2D\phi_x/\phi$$

and show that the auxiliary function $\phi(x, t)$ satisfies the diffusion equation, $\phi_t = D\phi_{xx}$. Thereby the Cole–Hopf transformation recasts the nonlinear Burgers equation into the linear diffusion equation.

(b) Consider an arbitrary initial condition $u(x, 0) = F(x)$. Show that

$$\phi(x, 0) = \Phi(x) = \exp\left[-(2D)^{-1}\int_0^x F(y)\,dy\right]$$

can be chosen as the initial condition for the diffusion equation.

(c) Starting with solution to the diffusion equation subject to the initial condition $\phi(x, 0) = \Phi(x)$,

$$\phi(x, t) = \frac{1}{\sqrt{4\pi Dt}}\int_{-\infty}^{\infty}\Phi(y)e^{-(x-y)^2/4Dt}\,dy,$$

show that the solution of the Burgers equation subject to $u(x, 0) = F(x)$ is

$$u(x, t) = \frac{\displaystyle\int_{-\infty}^{\infty}[(x-y)/t]\,e^{-G/2D}\,dy}{\displaystyle\int_{-\infty}^{\infty}e^{-G/2D}\,dy}, \qquad G(x, y; t) = \int_0^y F(y')\,dy' + \frac{(x-y)^2}{2t}.$$

4.7 Specialize the general solution described in the previous problem to the Burgers equation (4.34) with initial condition $\rho_0(x) = 1$ for $x \le 0$ and $\rho_0(x) = 0$ for $x > 0$. Show that in the bulk region $|x| < t$ the solution approaches $\rho = \frac{1}{2}(1-x/t)$. Describe the behavior in the internal layers around $x = \pm t$.

4.8 Consider an open system of size $N = 3$. Verify that the matrix ansatz gives the stationary solution.

4.9 Complete the proof of the validity of relation (4.54).

4.10 Consider the configuration $\underbrace{1\ldots1}_{M}\underbrace{0\ldots0}_{N-M}$ and show that its stationarity is equivalent to

$$\langle W|D^M E^{N-M}|V\rangle = \alpha\langle W|ED^{M-1}E^{N-M}|V\rangle + \beta\langle W|D^M E^{N-M-1}D|V\rangle.$$

Check the validity of the above relation using (4.52).

4.11 Verify the identity (4.57) on the line $\alpha + \beta = 1$ where we can replace the operators by scalars: $E = \alpha^{-1}, D = \beta^{-1}$. Then prove (4.57) in the general case using mathematical induction.

4.12 The derivation of the expressions for the current and density in different phases, Eqs (4.59)–(4.60), is lengthy as it requires the asymptotic analysis of various sums like the one that appears in (4.58). It is always useful to get an idea of what to expect

by considering the line $\alpha + \beta = 1$ where the solution greatly simplifies (since the operators E and D are scalar) and the major results are independent of N. Show that, on the line $\alpha + \beta = 1$, $\langle n_i \rangle = \alpha$ independent of both i and N. Verify that this result agrees with the general prediction of (4.60).

4.13 Consider the ASEP in an open system, with equal input and output rates, $\alpha = \beta$.

(a) Using l'Hôpital's rule simplify the sum in (4.58) to yield

$$\frac{\langle W|C^N|V\rangle}{\langle W|V\rangle} = \sum_{p=1}^{N} p \, \frac{(2N-1-p)!}{N!(N-p)!} \, \frac{p+1}{\alpha} .$$

(b) Show that when $\alpha > 1/2$, the terms with $p = \mathcal{O}(1)$ dominate. Use this fact to compute the leading asymptotic

$$\frac{\langle W|C^N|V\rangle}{\langle W|V\rangle} \simeq \frac{\alpha^2}{\sqrt{\pi}\,(2\alpha-1)^3} \frac{4^{N+1}}{N^{3/2}} .$$

(c) Show that for $\alpha < 1/2$, the terms with p close to $N(1-2\alpha)/(1-\alpha)$ dominate. Use this fact to compute the leading asymptotic

$$\frac{\langle W|C^N|V\rangle}{\langle W|V\rangle} \simeq \frac{(1-2\alpha)^2}{(1-\alpha)^2} \frac{N}{\alpha^N(1-\alpha)^N} .$$

(d) Show that when $\alpha = 1/2$,

$$\frac{\langle W|C^N|V\rangle}{\langle W|V\rangle} = 4^N .$$

(e) Use the results of (b)–(d) to confirm the predictions of (4.59) on the diagonal $\alpha = \beta$.

(f) Confirm the predictions of (4.60) on the diagonal $\alpha = \beta$.
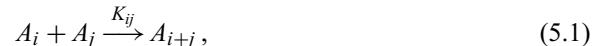
5   Aggregation

In aggregation, reactive clusters join irreversibly whenever two of them meet. Aggregation is ubiquitous in Nature: it underlies milk curdling, blood coagulation, and star formation by gravitational accretion. Aggregation also beautifully illustrates some paradigmatic features of non-equilibrium phenomena, such as scaling, phase transitions, and non-trivial steady states. Here we describe these generic features and present exact solutions of a few tractable models of aggregation.

5.1 The master equations

In kinetic theory, which was the starting point of non-equilibrium statistical physics, the dynamics involves a continuing sequence of scattering events: $P + P \rightarrow P + P$. In aggregation the primary process may be represented as

$$A_i + A_j \xrightarrow{K_{ij}} A_{i+j}, \tag{5.1}$$

in which two clusters of mass i and j irreversibly join at rate K_{ij} to form a cluster of mass $i+j$ (Fig. 5.1). Aggregation is manifestly irreversible, as the number of clusters decreases with time and eventually all clusters merge into a single cluster. In contrast to scattering processes, where the constituent particles keep their identities, here continuous merging occurs. To highlight this distinction with scattering, it is customary to refer to the aggregating entities as *clusters*. The goal of this chapter is to determine the time evolution of the concentration of clusters of mass k, $c_k(t)$, and to understand which features of the underlying reaction rate, or kernel, K_{ij}, determine this distribution.

Throughout this chapter we tacitly assume that the mass k runs over the positive integers – this merely implies that we measure mass in terms of a minimal mass, and a cluster of mass k contains k such primal, minimal-mass clusters. These elemental clusters are called
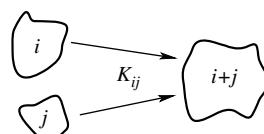
Fig. 5.1.   Clusters of mass *i* and mass *j* merge irreversibly into a cluster of mass $i + j$.

monomers, while clusters of mass k are termed k-mers. With this convention for the mass, the reaction rates form an infinite symmetric matrix $K_{i,j} = K_{j,i}$.

The starting point for treating aggregation is an infinite set of master equations that describe how the cluster mass distribution $c_k(t)$ evolves. These equations are

$$\frac{dc_k}{dt} = \frac{1}{2} \sum_{i+j=k} K_{ij}\, c_i\, c_j - c_k \sum_{i \geq 1} K_{ik}\, c_i\,. \tag{5.2}$$

The first term on the right-hand side accounts for the creation of a k-mer due to the coalescence of two clusters of masses i and j with $i + j = k$. The prefactor $\frac{1}{2}$ in the gain term is needed to avoid double counting. For example, when $k = 5$ we must sum over $(i,j) = (1,4), (2,3), (3,2), (4,1)$ and therefore the gain term is $K_{1,4}\, c_1 c_4 + K_{2,3}\, c_2 c_3$ in accordance with the definition of the reaction rate matrix (using symmetry $K_{i,j} = K_{j,i}$). For clusters of even mass the prefactor $\frac{1}{2}$ remains in one of the gain terms; e.g. for clusters of mass $k = 6$ the gain term is

$$K_{1,5}\, c_1 c_5 + K_{2,4}\, c_2 c_4 + \frac{1}{2}\, K_{3,3}\, (c_3)^2\,.$$

To understand this subtlety in accounting for reactions between same-mass clusters, consider a finite system. Let N_k be the total number of clusters of mass k. For $i \neq j$ there are $N_i N_j$ pairs of type ij, while the number of same-mass pairs is $\frac{1}{2} N_k (N_k - 1) \to \frac{1}{2} N_k^2$ in the thermodynamic limit. Thus the prefactor $\frac{1}{2}$ correctly accounts for the relative fraction of same-mass pairs.

The loss term on the right-hand side of Eq. (5.2) accounts for reactions of a k-mer with other clusters. The loss term for same-mass pairs is $K_{jj} c_j c_j$ rather than $\frac{1}{2} K_{jj} c_j c_j$ since *two* clusters of mass j disappear in such a collision. A foolproof way to account for this relative factor of $\frac{1}{2}$ is to write Eq. (5.2) as

$$\frac{dc_k}{dt} = \frac{1}{2} \sum_{i,j \geq 1} K_{ij}\, c_i c_j\, [\delta_{i+j,k} - \delta_{i,k} - \delta_{j,k}],$$

where $\delta_{i,j}$ is the Kronecker delta function, so that all the ways that a k-mer can be created or lost are counted in a manifestly correct way.

The master equations are the starting point for almost all studies of aggregation, and it is important to keep in mind the assumptions underlying this approach:

- *Cluster locations*. Most importantly, the locations of clusters are ignored. Namely, the system is assumed to be well mixed so that the probability that multiple reactants are in close proximity factorizes as the product of single-particle densities – this is the *mean-field* approximation.
- *Bimolecular reactions*. The system is sufficiently dilute that higher-body interactions are negligible.
- *Shape independence*. The aggregate mass is the only dynamical variable; cluster shape plays no role in the evolution (as in the aggregation of spherical liquid droplets).

- *Thermodynamic limit*. The system is sufficiently large that cluster concentrations are continuous functions; discreteness effects are ignored.

The mean-field approximation is often asymptotically exact in the physical case of three dimensions. In Chapter 13 we shall discuss how to deal with lower-dimensional systems where the mean-field approximation does not apply. The restriction to bimolecular reactions is not serious because the cluster density decreases in irreversible aggregation. Thus a system that is initially not dilute eventually becomes dilute, and multiple collisions play an asymptotically negligible role.[1] The assumption of shape independence is hard to justify theoretically, although it has been validated numerically for a specific aggregation model. In many cases, however, clusters are spherical because of additional physical mechanisms such as surface tension. The thermodynamic limit is applicable for large systems.

We stress that within the master equation approach the reaction rates are assumed given, while in reality the determination of the rates may be part of the analysis. The main feature of the reaction that influences the rates is the transport mechanism. Most applications of aggregation are in physical chemistry where transport is caused by diffusion. Other transport mechanisms are certainly possible; for example, ballistic transport is relevant for astrophysical applications.[2] Equations such as (5.2) have also been used in ballistically controlled aggregation. In this case, the validity of (5.2) is difficult to justify since each cluster is characterized by its mass and velocity, and the latter should enter in a master equation description. In such systems, the mean-field framework now involves an infinite set of Boltzmann-like equations.[3]

Equations (5.2) admit a basic integral of motion – the mass density

$$M(t) = \sum_{k \geq 1} k \, c_k(t) \tag{5.3}$$

is conserved. To verify this conservation law we write

$$\frac{dM}{dt} = \sum_k k \frac{dc_k}{dt} = \sum_k \sum_{i+j=k} \frac{1}{2} K_{ij} (i+j) \, c_i \, c_j - \sum_k \sum_i K_{ik} \, k \, c_i \, c_k = 0. \tag{5.4}$$

The outer sum over k causes the sums over i and j in the gain term to become independent and unrestricted. Thus the gain and loss terms cancel and the mass density is manifestly conserved. As we shall see, mass conservation plays a central role in the description of the cluster mass distribution.

---

[1] There are exceptions to this rule, however; for example, many natural processes involve higher-body interactions in an essential way, such as catalyst-facilitated reactions. Moreover, some reactions are isomorphic to aggregation with multi-cluster merging. Later in this book we shall investigate reaction processes that evolve by three-body merging: $A_i + A_j + A_k \rightarrow A_{i+j+k}$ (see Sections 8.8, 9.7, and 10.2).

[2] Occasionally, more exotic transport mechanisms can occur. For instance, in turbulence the average distance between particles grows as $t^{3/2}$, so that the transport is superballistic.

[3] See our discussions of ballistic agglomeration and traffic in Sections 3.7 and 3.8.

5.2 Exact solution methods

The master equations are a formidable infinite set of coupled nonlinear differential equations that are soluble only for a few special kernels. Many clever solution techniques have been developed for these kernels that we now present. We frame our discussion in the framework of the constant reaction kernel, proposed and solved by Smoluchowski in 1917, because its simplicity provides an ideal playground to illustrate many exact approaches.

To justify the constant-kernel model, let's start with the reaction kernel for the aggregation of clusters that move by Brownian motion. The rate at which particles of diffusivity D hit a sphere of radius R is proportional to DR^{d-2}, or DR in three dimensions (see Section 2.7). The generalization to a collection of diffusing spherical aggregates, each of mass i, radius R_i, and diffusivity D_i, becomes $K_{ij} \sim (D_i + D_j)(R_i + R_j)$. In this formula, the radius of a cluster scales as the cube root of its mass, while the diffusivity is proportional to the inverse radius. The latter behavior is a consequence of the Stokes–Einstein relation between the diffusivity D of a spherical object and its drag coefficient $6\pi \eta R$; namely, $D = T/(6\pi \eta R)$ (as discussed in Section 2.8). As a result, the reaction rate becomes

$$K_{ij} \sim (i^{-1/3} + j^{-1/3})(i^{1/3} + j^{1/3}) = 2 + \left(\frac{i}{j}\right)^{1/3} + \left(\frac{j}{i}\right)^{1/3}. \qquad (5.5)$$

The master equations (5.2) with this Brownian kernel are still unsolved. However, the Brownian kernel shares one important feature with the constant kernel – both are invariant under the transformation $(i,j) \longrightarrow (ai, aj)$, that is, $K_{ai,aj} = K_{i,j}$. This feature suggests that the constant kernel is a reasonable (albeit uncontrolled) approximation for the physically important Brownian kernel.

For the constant kernel, we choose $K_{ij} = 2$ for convenience, and then the master equations become

$$\frac{dc_k}{dt} = \sum_{i+j=k} c_i c_j - 2c_k \sum_{i \geq 1} c_i \equiv \sum_{i+j=k} c_i c_j - 2c_k N, \qquad (5.6)$$

where $N(t) = \sum_{k \geq 1} c_k(t)$ is the concentration of clusters of any mass. The first few of these equations are

$$\begin{aligned}
\dot{c}_1 &= -2c_1 N, \\
\dot{c}_2 &= c_1^2 - 2c_2 N, \\
\dot{c}_3 &= 2c_1 c_2 - 2c_3 N, \\
\dot{c}_4 &= 2c_1 c_3 + c_2^2 - 2c_4 N, \\
\dot{c}_5 &= 2c_1 c_4 + 2c_2 c_3 - 2c_5 N, \\
\dot{c}_6 &= 2c_1 c_5 + 2c_2 c_4 + c_3^2 - 2c_6 N,
\end{aligned} \qquad (5.7)$$

where the overdot denotes the time derivative.

One major lesson that emerges from studies of aggregation and other irreversible processes is that the long-time behavior (which is the most interesting characteristic of the system) depends on the initial condition in a trivial way – through the total mass density – while details of the initial data are irrelevant. Therefore it is natural to choose the monomer-only initial condition

$$c_k(0) = \delta_{k,1} . \tag{5.8}$$

If not stated otherwise, we shall always assume such an initial condition in the following.

We now outline several methods of solution for constant-kernel aggregation to provide a sense of the range of techniques to solve master equations of the form (5.6).

Moments

Before solving the initial value problem (5.6)–(5.8), it is instructive to study the moments of the mass distribution,

$$M_n(t) \equiv \sum_{k \geq 1} k^n c_k(t) , \tag{5.9}$$

where considerable information can be gleaned with little effort, even in cases where the full master equations are unsolvable. Moments also give basic information about the mass distribution itself; for example, the ratio $M_1/M_0 \equiv M/N$ gives the average cluster mass.

For constant reaction rates, the moment equations are particularly simple. Using Eqs (5.6) we deduce

$$\frac{dM_n}{dt} = \sum_{k \geq 1} k^n \frac{dc_k}{dt} = \sum_{k \geq 1} k^n \left[\sum_{i+j=k} c_i c_j - 2 c_k \sum_{i \geq 1} c_i \right]$$

$$= \sum_{i,j} (i+j)^n c_i c_j - 2 M_n M_0, \tag{5.10}$$

where the sums over i and j are unrestricted in the second line. The explicit equations for the first few moments are

$$\dot{M}_0 = \sum_{i,j} c_i c_j - 2 M_0^2 = -M_0^2,$$

$$\dot{M}_1 = \sum_{i,j} (i+j) c_i c_j - 2 M_1 M_0 = 0,$$

$$\dot{M}_2 = \sum_{i,j} (i^2 + 2ij + j^2) c_i c_j - 2 M_2 M_0 = 2 M_1^2, \tag{5.11}$$

$$\dot{M}_3 = \sum_{i,j} (i^3 + 3i^2 j + 3ij^2 + j^3) c_i c_j - 2 M_3 M_0 = 6 M_1 M_2.$$

For the monomer-only initial condition, $M_n(0) = 1$ for $n \geq 0$. Then the solution for the zeroth moment $M_0 = N$ is

$$N(t) = \frac{1}{1+t}. \tag{5.12}$$

Solving Eqs (5.11) for the higher moments one by one we obtain $M_1 = 1$, $M_2 = 1 + 2t$, $M_3 = 1 + 6t + 6t^2$, etc. In general, $M_n \simeq n! \, t^{n-1}$ as $t \to \infty$. Note that the total mass is conserved, $M_1 = 1$, as it must be. A natural measure of the typical cluster mass is provided by the second moment M_2, which grows linearly with time.

Recursive approach

Equations (5.6) are recursive and therefore can be solved one by one. For the monomer-only initial condition, we substitute $N(t)$ from (5.12) into the first of (5.7) and integrate to give $c_1(t) = (1+t)^{-2}$. Having found c_1, the equation for c_2 becomes

$$\frac{dc_2}{dt} = (1+t)^{-4} - 2(1+t)^{-1} c_2.$$

Solving this equation subject to $c_2(0) = 0$ gives $c_2(t) = t/(1+t)^3$. The next density satisfies

$$\frac{dc_3}{dt} = 2t(1+t)^{-5} - 2(1+t)^{-1} c_3,$$

whose solution, with $c_3(0) = 0$, is $c_3(t) = t^2/(1+t)^4$. Continuing this recursive approach we find $c_4(t) = t^3/(1+t)^5$, then $c_5(t) = t^4/(1+t)^6$, etc. This pattern suggests the general solution

$$c_k(t) = \frac{t^{k-1}}{(1+t)^{k+1}}. \tag{5.13}$$

Using induction one may readily verify that this guess is correct.

The elegant closed-form solution (5.13) has many important properties, including the following:

1. For $k, t \to \infty$, $c_k \to t^{-2} e^{-k/t}$. Thus for fixed large k and $t \to \infty$, each $c_k(t)$ approaches the common limit t^{-2}, while for $k < t$, the mass distribution is nearly flat (Fig. 5.2).
2. The integral of cluster concentration, $\int c_k \, dk$, is proportional to $t^{-2} \times t = t^{-1}$. This reproduces the correct time dependence for the total concentration of clusters.
3. For an arbitrary initial mass distribution with finite support[4] the short- and long-time limits of c_k can be easily determined without solving the full master equations. For the short-time behavior, we ignore the loss terms and obtain equations of the form in Eqs (5.14) below, for which $c_k(t) \sim t^{k-1}$ for $t \ll 1$. Conversely for $t \to \infty$, there is no production of k-mers for fixed k. We therefore ignore the gain terms in the master equation to give $\dot{c}_k \sim -2c_k N$, whose solution is $c_k \sim t^{-2}$.

---

[4] The distribution $c_k(0)$ has a finite support if $c_k(0) \equiv 0$ for all $k > k_{max}$, with finite k_{max}.
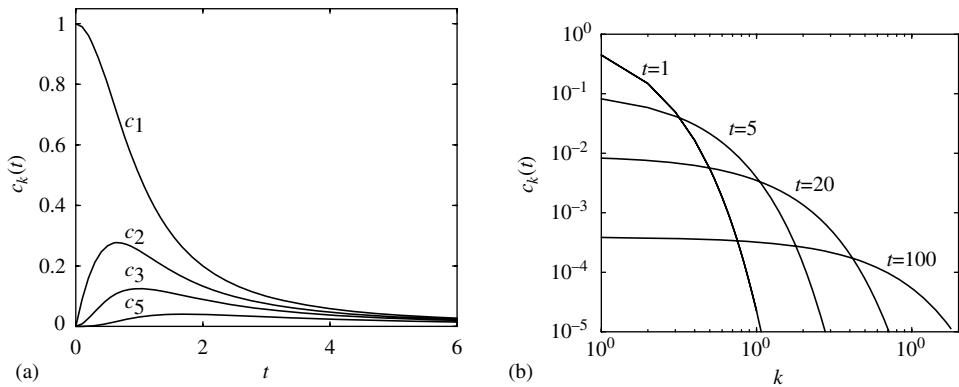
Fig. 5.2. (a) Cluster concentrations $c_k(t)$ versus time for constant-kernel aggregation for $k = 1, 2, 3,$ and 5. (b) Cluster concentrations $c_k(t)$ versus k on a double-logarithmic scale for $t = 1, 5, 20,$ and 100.

Elimination of loss terms

One useful trick that often simplifies master equations is to eliminate the loss terms by considering concentration ratios, rather than the concentrations themselves. For example, consider the concentration ratio $\phi_k \equiv c_k/c_1$, whose evolution equation is readily shown to be

$$\frac{d\phi_k}{dt} = c_1 \sum_{i+j=k} \phi_i \phi_j \,. \tag{5.14}$$

Thus the loss term has indeed disappeared. We can also eliminate the prefactor on the right-hand side by introducing an artificial time

$$\tau = \int_0^t dt' \, c_1(t'), \tag{5.15}$$

so that Eq. (5.14) reduces to

$$\frac{d\phi_k}{d\tau} = \sum_{i+j=k} \phi_i \phi_j \,. \tag{5.16}$$

To relate τ and the time t, we substitute the previously established monomer density $c_1 = (1+t)^{-2}$ into (5.15) and find $\tau = t/(1+t)$. Solving (5.16) for the first few ϕ_k, we get $\phi_1 = 1, \phi_2 = \tau, \phi_3 = \tau^2$; one can easily check that the general solution is $\phi_k = \tau^{k-1}$. Finally, substituting into $c_k = \phi_k \, c_1$ and using $\tau = t/(1+t)$, we rederive (5.13).

Exponential ansatz

Solutions to master equations often have an exponential form – Eq. (5.13) is one such example. By using this assumption at the outset, we can simplify the governing equations

considerably. For the case of constant-kernel aggregation, the appropriate exponential ansatz is

$$c_k(t) = A\, a^{k-1}, \tag{5.17}$$

with as yet unknown $A(t)$ and $a(t)$. The initial condition (5.8) implies $A(0) = 1$ and $a(0) = 0$; this fact explains why we choose the power $k - 1$ rather than k.

Substituting ansatz (5.17) into Eq. (5.6), and dividing both sides of the equation by c_k, gives

$$\frac{\dot{A}}{A} + (k-1)\frac{\dot{a}}{a} = (k-1)\frac{A}{a} - \frac{2A}{1-a}\,.$$

The k-dependent and k-independent components of this equation should separately vanish to ensure the validity of the exponential ansatz. Hence the problem reduces to solving the coupled differential equations

$$\dot{A} = -\frac{2A^2}{1-a}, \qquad \dot{a} = A. \tag{5.18}$$

If we had chosen a different power of a in the initial ansatz, there would not be the alignment of k-dependent terms given above, but it would also be clear from the degree of misalignment how to choose the correct power of a. Since $\sum_{k\geq1} kc_k = A\sum_{k\geq1} ka^{k-1} = A(1-a)^{-2}$, mass conservation implies $A = (1-a)^2$. Substituting $A = (1-a)^2$ into (5.18) we immediately find

$$A = \frac{1}{(1+t)^2}, \qquad a = \frac{t}{1+t}, \tag{5.19}$$

thus reproducing the solution for $c_k(t)$ in Eq. (5.13).

The exponential ansatz has an advantage over the two previous approaches in that it involves less guesswork and it requires dealing with two (instead of infinitely many) differential equations. This ansatz also works for all initial conditions that decay exponentially with mass.

Example 5.1. *Exponentially decaying initial data.* Let $c_k(0) = (1-q)^2 q^{k-1}$ where $0 < q < 1$; the monomer-only initial condition is recovered for $q \to 0$. We use the exponential ansatz (5.17) with $A = (1-a)^2$ and now we must solve $\dot{a} = (1-a)^2$ subject to $a(0) = q$. We find

$$a(t) = 1 - \frac{1-q}{1+(1-q)t}, \qquad A(t) = \left[\frac{1-q}{1+(1-q)t}\right]^2.$$

The parameter q does not affect the qualitative asymptotic behavior. For instance, for fixed mass, the k-mer concentrations approach a common limit that scales as t^{-2} as $t \to \infty$. When both mass and time diverge in such a way that the ratio k/t remains finite, the mass distribution again attains the scaling form $c_k \simeq t^{-2}\exp(-k/t)$.

Generating function method

A powerful approach for solving for the cluster mass distribution is the generating function method. This technique is ideally suited for aggregation because the master equations are in the form of a discrete convolution. Using the generating function, this convolution is transformed into a readily soluble product. The generating function is defined as

$$\mathcal{C}(z, t) \equiv \sum_{k \geq 1} c_k(t) \, z^k, \tag{5.20}$$

and it encodes the entire mass distribution within a single function. Notice that the generating function is simply the discrete analog of the Laplace transform. Thus insights about the relation between a function and its Laplace transform carry over to the generating function.

To apply the generating function method to constant-kernel aggregation, we multiply Eqs (5.6) by z^k, and sum over all k. This gives

$$\frac{d\mathcal{C}}{dt} = \sum_{k \geq 1} \sum_{i+j=k} c_i z^i \, c_j z^j - 2 \sum_{k \geq 1} c_k z^k \sum_{i \geq 1} c_i = \mathcal{C}^2 - 2\,\mathcal{C}N. \tag{5.21}$$

Here the sum over k renders the two sums over i and j independent, so that the first term reduces to a product. This reduction is the essential simplification of the generating function method. We can make an additional simplification by noting that the equation of motion for N is $\dot{N} = -N^2$ so that the function $\mathcal{C} - N$ satisfies $d(\mathcal{C} - N)/dt = (\mathcal{C} - N)^2$ – we've reduced the entire set of master equations to a single differential equation! This equation should be supplemented with the initial condition. If only monomers exist initially, then $\mathcal{C}(z, t = 0) = z$. Solving gives $\mathcal{C} - N = (z - 1)/[1 - (z - 1)t]$, from which

$$\mathcal{C} = \frac{1}{1 + t} \frac{z}{1 - (z - 1)t} . \tag{5.22}$$

Rearranging (5.22) slightly and expanding in a power series in z gives

$$\mathcal{C}(z, t) = \sum_{k \geq 1} z^k \frac{t^{k-1}}{(1 + t)^{k+1}} .$$

From this form, we directly read off the mass distribution and thereby recover Eq. (5.13).

For an *arbitrary* initial condition the generating function is

$$\mathcal{C}(z, t) = (1 + t)^{-2} \frac{\mathcal{C}_0(z)}{1 - [t/(1 + t)]\,\mathcal{C}_0(z)} , \tag{5.23}$$

where $\mathcal{C}_0(z) = \mathcal{C}(z, t = 0)$ and we also assume that $N(t = 0) = \mathcal{C}_0(z = 1) = 1$. Expanding the generating function (5.23) as a power series in z to obtain the densities $c_k(t)$ for all k is straightforward in principle, but can become computationally tedious.

Example 5.2. *Initial mass distribution with finite support.* Consider an initial state that consists of a mixture of monomers and dimers, that is $c_1(0), c_2(0) > 0$, with $c_1(0) + c_2(0) =$

1, while $c_j(0) = 0$ for $j \geq 3$. Then $\mathcal{C}_0(z) = c_1(0)z + c_2(0)z^2$ so that (5.23) gives

$$\mathcal{C}(z,t) = (1+t)^{-2} \; \frac{c_1(0)z + c_2(0)z^2}{1 - [t/(1+t)][c_1(0)z + c_2(0)z^2]}.$$

To expand this generating function in powers of z, we rewrite the quadratic polynomial in the denominator as $(1 - z/z_1)(1 - z/z_2)$, where $z_{1,2}$ are the roots of the polynomial. Then we represent $[(1 - z/z_1)(1 - z/z_2)]^{-1}$ as a combination of $(1 - z/z_1)^{-1}$ and $(1 - z/z_2)^{-1}$, and expand each of these factors as a geometric series.

This same approach allows us to derive explicit results if $c_k(0) = 0$ for $k \geq 5$, where the fourth-order polynomial in the denominator in (5.23) can be explicitly factorized. More generally, the generating function solution allows us to deduce the *asymptotic* behavior of the cluster concentrations $c_k(t)$ for arbitrary initial mass distributions with finite support. The result of this inversion procedure is (see problem 5.1)

$$c_k \simeq \frac{1}{M(0)\, t^2} \; e^{-k/[M(0)t]} \,. \tag{5.24}$$

Inversion of generating functions

Often, we are able to solve for the generating function $\mathcal{C}(z) = \sum_{k \geq 1} c_k\, z^k$, and we want to extract the large-k asymptotic properties of an underlying mass distribution c_k. We now show how to relate the large-k tail of c_k with the behavior of $\mathcal{C}(z)$ for $z \to 1$. As a generic example, suppose that

$$c_k \simeq C k^{-\alpha} \qquad \text{for} \quad k \gg 1. \tag{5.25}$$

We restrict ourselves to the regime $\alpha > 1$, so that $\sum_k c_k$ converges; we further set $\sum_{k \geq 1} c_k = 1$. For now, suppose that α is not equal to 2. Then in the limit $z \uparrow 1$, the generating function $\mathcal{C}(z)$ and c_k are asymptotically related by

$$\mathcal{C}(z) = 1 + C\Gamma(1 - \alpha)\,(1 - z)^{\alpha - 1} + \cdots, \tag{5.26}$$

where Γ is the Euler gamma function.

To establish this result, first note that $\mathcal{C}(z = 1) = 1$ by hypothesis. Next we differentiate $\mathcal{C}(z)$ with respect to z, use (5.25) for c_k, and take the $z \uparrow 1$ limit to yield

$$\frac{d\mathcal{C}}{dz} = \sum_{k \geq 1} k\, c_k\, z^{k-1} \simeq C \int_0^\infty dk\, k^{1-\alpha}\, e^{k \ln z}$$

$$\simeq C \int_0^\infty dk\, k^{1-\alpha}\, e^{-k(1-z)}$$

$$= C\,\Gamma(2 - \alpha)\,(1 - z)^{\alpha - 2} \,. \tag{5.27}$$

Integrating this equation gives (5.26). We previously used this same trick of working with the more singular function $d\mathcal{C}/dz$ rather than $\mathcal{C}$ itself (see Section 2.5) because this device accentuates the relevant asymptotic behavior.

In the marginal case of $\alpha = 2$, the same steps in Eq. (5.27) (that led to Eq. (5.26)) yield

$$\mathcal{C}_0(z) = 1 + C\,(1-z)\,\ln(1-z) + \cdots . \tag{5.28}$$

Example 5.3. *Algebraically decaying initial data*. If the initial mass distribution is unbounded and decays sufficiently slowly, anomalous behavior of the cluster mass distribution may occur. One such example is given by the initial distribution

$$c_k(0) \to C\,k^{-\alpha} \tag{5.29}$$

for $k \gg 1$. The exponent α should be larger than one, since the zeroth moment of the mass distribution must converge – otherwise Eqs (5.6) are ill-defined. Apart from this restriction, the decay exponent α is arbitrary. In particular, if $\alpha \le 2$, the first moment diverges, $M(0) = \infty$, and the asymptotic form given in Eq. (5.24) no longer holds.

Suppose that $1 < \alpha < 2$. Then from Eq. (5.25) the generating function $\mathcal{C}_0(z)$ that encodes the initial data has the following asymptotic behavior in the $z \uparrow 1$ limit:

$$\mathcal{C}_0(z) = 1 + C\,\Gamma(1-\alpha)\,(1-z)^{\alpha-1} + \cdots . \tag{5.30}$$

Using (5.30), taking the limits $t \to \infty$ and $z \uparrow 1$, and keeping only the leading term, the generating function (5.23) becomes

$$\mathcal{C}(z,t) = \frac{t^{-1}}{1 + Dt(1-z)^{\alpha-1}}, \qquad D = -C\,\Gamma(1-\alpha).$$

From the grouping of variables in the denominator, we identify the scaling variable as $w = k/(Dt)^{1/(\alpha-1)}$, and

$$c_k \simeq t^{-1}(Dt)^{-1/(\alpha-1)} f_\alpha(w), \quad w = k/(Dt)^{1/(\alpha-1)}, \tag{5.31}$$

as the scaling form of the mass distribution. Indeed,

$$\mathcal{C}(z,t) = \sum_{k \ge 1} c_k(t)z^k \simeq t^{-1} \int_0^\infty dw\, f_\alpha(w)\, e^{-ws} = \frac{t^{-1}}{1 + s^{\alpha-1}},$$

where $s = (Dt)^{1/(\alpha-1)}(1-z)$. Hence the scaling function $f_\alpha(w)$ is the inverse Laplace transform of $(1 + s^{\alpha-1})^{-1}$.

The small- and large-s asymptotics of the Laplace transform imply the corresponding large- and small-w behaviors of the scaling function:

$$f_\alpha(w) \simeq \begin{cases} \dfrac{1}{\Gamma(\alpha - 1)} \, w^{-(2-\alpha)}, & w \to 0, \\[3mm] -\dfrac{1}{\Gamma(1 - \alpha)} \, w^{-\alpha}, & w \to \infty. \end{cases} \tag{5.32}$$

For certain values of α, it is possible to obtain an explicit expression for the scaling function. For instance when $\alpha = 3/2$, the inverse Laplace transform of $(1 + \sqrt{s})^{-1}$ admits the representation in terms of the error function

$$f_{3/2}(w) = \frac{1}{\sqrt{\pi w}} - e^w \operatorname{erfc}\left(\sqrt{w}\right).$$

In the marginal case $\alpha = 2$, the scaling behavior is simple. Now we use Eq. (5.28) to find that the scaling solution is still an exponential function

$$c_k(t) = t^{-2}(C \ln t)^{-1} \exp\left[-\frac{k}{Ct \ln t}\right],$$
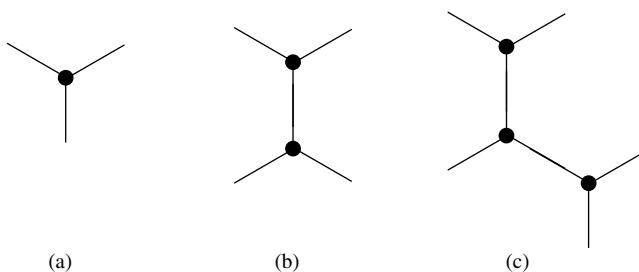
but with logarithms appearing in the scaling variable and in the prefactor. These logarithmic factors are the only features that distinguish the behavior in the marginal case from the standard scaling behavior for all $\alpha > 2$. For $\alpha \geq 2$ the scaling function is universal and purely exponential, while for $1 < \alpha < 2$ the scaling function $f_\alpha(w)$ is the inverse Laplace transform of $(1 + s^{\alpha-1})^{-1}$. This scaling function is still universal in that it is independent of details of the initial data (small-mass behavior, the amplitude C, etc.); the only relevant feature is the magnitude of the decay exponent α.

5.3 Gelation

When the aggregation rate is a sufficiently increasing function of the masses of the reacting clusters, *gelation* can occur in a finite time in which a non-zero fraction of the total mass condenses into a single cluster (think of gelatin dessert, or "Jello"). As we now discuss, the product kernel represents an exactly soluble example of this spectacular feature. Beyond the gelation time, the system divides into two phases: the *gel*, or the infinite cluster, and the remaining *sol* of finite clusters whose total mass decreases with time because the remaining finite clusters continue to merge with the gel.

Product kernel, $K_{ij} = ij$

The product kernel arises when monomers consist of f-functional reactive endgroups (Fig. 5.3). When two such monomers merge, the resulting dimer has $2f - 2$ reactive endgroups, a trimer has $3f - 4$ endgroups, and a k-mer has $kf - 2(k - 1) = (f - 2)k + 2$

Fig. 5.3. Small k-mers of 3-functional units: (a) monomer, (b) dimer, and (c) trimer.

endgroups. If all endgroups are equally reactive, the reaction rate between two clusters equals the product of the number of endgroups:

$$K_{ij} = [(f-2)i+2][(f-2)j+2] = (f-2)^2 ij + 2(f-2)(i+j) + 4. \qquad (5.33)$$

The case $f = 2$ corresponds to linear polymers, for which K_{ij} is constant, while the product kernel arises for $f \to \infty$. For finite $f > 2$, the kernel is a linear combination of the constant, product, and sum kernels.

Let us now focus on the pure product kernel, $K_{ij} = ij$, for which the master equations are[5]

$$\frac{dc_k}{dt} = \frac{1}{2} \sum_{i+j=k} ij\, c_i c_j - kc_k \sum_{i \geq 1} i\, c_i = \frac{1}{2} \sum_{i+j=k} ij\, c_i c_j - kc_k. \qquad (5.34)$$

An amazing feature of this infinite system of equations is that a finite fraction of all the mass condenses into a single giant cluster – the gel cluster – that eventually engulfs the entire system. Mathematically, the appearance of the gel is manifested by a singularity.

Moments

Let us try to detect this singularity in the simplest possible way by examining the moments of the mass distribution. Consider a finite system with initial mass $\mathcal{M}$. For such a system, the formation of an "infinite" gel molecule means the appearance of a cluster whose mass $g\mathcal{M}$ is a finite fraction of the total mass $\mathcal{M}$, and whose concentration is $1/\mathcal{M}$. We now decompose the moments of the mass distribution,

$$M_n = \sum_{k \geq 1} k^n\, c_k = \sum_{\text{sol}} k^n\, c_k + (k^n c_k)_{\text{gel}}\,,$$

---

[5] Product-kernel aggregation is equivalent to the Erdős–Rényi random graph that will be discussed in Section 14.2. Thus the solution to product-kernel aggregation provides many fundamental properties of the Erdős–Rényi random graph for free.

as a sum $\sum_{\text{sol}}$ over finite clusters – the sol – and the single "infinite" gel cluster. According to this decomposition, the first few moments are[6]

$$M_0 = \sum_{\text{sol}} c_k,$$

$$M_1 = \sum_{\text{sol}} k c_k + g,$$

$$M_2 = \sum_{\text{sol}} k^2 c_k + g^2 \mathcal{M},$$

$$M_3 = \sum_{\text{sol}} k^3 c_k + g^3 \mathcal{M}^2.$$

In the thermodynamic $\mathcal{M} \to \infty$ limit, the second and higher moments therefore diverge, while prior to gelation, $t < t_g$, $g(t) \equiv 0$ and all moments are finite. This argument suggests examining the evolution of the second moment to detect the singularity.

From Eqs (5.34), the second moment M_2 evolves as

$$\frac{dM_2}{dt} = \sum_{k \geq 1} k^2 \frac{dc_k}{dt} = \frac{1}{2} \sum_{i \geq 1} \sum_{j \geq 1} (i+j)^2 (ic_i)(jc_j) - \sum_{k \geq 1} k^3 c_k$$

$$= \sum_{i \geq 1} \sum_{j \geq 1} (i^2 c_i)(j^2 c_j) = M_2^2 . \tag{5.35}$$

Solving this equation subject to $M_2(0) = 1$ we obtain $M_2(t) = (1-t)^{-1}$. This singularity is the sign of gelation, and it leads to the conjecture that gelation occurs at $t_g = 1$. Is it possible that the third moment diverges earlier, so that gelation must have occurred earlier? The answer is *no*. Indeed, writing the evolution equation for the third moment,

$$\frac{dM_3}{dt} = \frac{1}{2} \sum_{i \geq 1} \sum_{j \geq 1} (i+j)^3 ic_i jc_j - \sum_{k \geq 1} k^4 c_k = 3M_3 M_2 = \frac{3}{1-t} M_3 , \tag{5.36}$$

and solving subject to $M_3(0) = 1$, we obtain $M_3(t) = (1-t)^{-3}$. Similarly,

$$\frac{dM_4}{dt} = \frac{1}{2} \sum_{i \geq 1} \sum_{j \geq 1} (i+j)^4 ic_i jc_j - \sum_{k \geq 1} k^5 c_k = 4M_4 M_2 + 3M_3^2 , \tag{5.37}$$

whose solution is $M_4(t) = (1 + 2t)(1 - t)^{-5}$. Using induction one may verify that all moments diverge at $t_g = 1$. Note that the location of the singularity $t_g = 1$ is not universal, but depends on the initial condition, more precisely, on $M_2(0)$. For a general initial condition, the solution of (5.35) becomes singular at $t_g = 1/M_2(0)$, but the character of the singularity remains the same as that for the monodisperse initial condition.

---

[6] For the zeroth moment, the distinction between the density of all clusters and finite clusters is immaterial since there is only a single infinite cluster.

It is now instructive to study the zeroth moment. Summing Eqs (5.34) we find that $M_0 \equiv N$ obeys

$$\frac{dN}{dt} = \frac{1}{2} \sum_{i \geq 1} \sum_{j \geq 1} i c_i \, j c_j - \sum_{k \geq 1} k c_k = \frac{1}{2} - 1 = -\frac{1}{2} \, . \tag{5.38}$$

The solution $N(t) = 1 - \frac{1}{2}t$ vanishes at $t = 2$ and becomes negative when $t > 2$. This pathology is a consequence of using the relation $\sum_{k \geq 1} k c_k = 1$ above the gel point. This sum actually refers to finite clusters, so we should write $\sum_{k \geq 1} k c_k = 1 - g$. Consequently,

$$\frac{dN}{dt} = \frac{1}{2} (1 - g)^2 - (1 - g) = \frac{g^2 - 1}{2}, \tag{5.39}$$

when $t > t_g$, so that the number of clusters stops decreasing only at infinite time, where the gel fraction reaches 1.

To probe the behavior of the moments (and the mass distribution) beyond the gel point, we need a more complete analysis that is provided by the generating function.

Generating function approach

To solve Eqs (5.34), it is convenient to use the *exponential* generating function $\mathcal{E}(y, t) \equiv \sum_{k \geq 1} k \, c_k(t) \, e^{yk}$. This generating function encodes the sequence $k c_k$ instead of the sequence c_k and simplifies the ensuing analysis. To determine the governing equation for $\mathcal{E}$ we multiply the master equation for each $\dot{c}_k$ by $k \, e^{yk}$ and sum over k to obtain

$$\begin{aligned}
\frac{\partial \mathcal{E}}{\partial t} &= \frac{1}{2} \sum_{i \geq 1} \sum_{j \geq 1} (i + j) ij \, c_i c_j \, e^{yk} - \sum_{k \geq 1} k^2 c_k \, e^{yk} \\
&= \frac{1}{2} \sum_{i \geq 1} i^2 c_i \, e^{yi} \sum_{j \geq 1} j c_j \, e^{yj} + \frac{1}{2} \sum_{i \geq 1} i c_i \, e^{yi} \sum_{j \geq 1} j^2 c_j \, e^{yj} - \sum_{k \geq 1} k^2 c_k \, e^{yk} \\
&= (\mathcal{E} - 1) \frac{\partial \mathcal{E}}{\partial y}. \tag{5.40}
\end{aligned}$$

We obtain the inviscid Burgers equation – the simplest nonlinear hyperbolic equation – that was the focus of Section 4.3 on the hydrodynamic description of exclusion processes. The salient feature of the Burgers equation is the appearance of a shock wave; this feature directly corresponds to the appearance of a gel in product-kernel aggregation.

To solve the Burgers equation (5.40) we use the method of characteristics that was previously discussed in Section 4.3. For Eq. (5.40), the characteristics are the curves in the y–t plane with slope $y_t = 1 - \mathcal{E}$. Along the characteristic the generating function $\mathcal{E}$ is constant, and therefore $y = (1 - \mathcal{E})t + f(\mathcal{E})$, where $f(\mathcal{E})$ is determined from the initial condition. For the monomer-only initial condition, the initial generating function is $\mathcal{E}(t = 0) = \sum k c_k \, e^{yk} \big|_{t=0} = e^y$, or $y(t = 0) = f(\mathcal{E}) = \ln \mathcal{E}$. We therefore arrive at the implicit solution

$$\mathcal{E} \, e^{-\mathcal{E}t} = e^{y-t}. \tag{5.41}$$

<div style="background:#ddd;padding:4px">Lagrange inversion</div>

Given a function $X = f(Y)$, with $X \simeq Y$ for small Y, what is the power series representation of the inverse function $Y(X) = \sum_{n \geq 1} A_n X^n$? The coefficients A_n are given by the *Lagrange inversion formula*. Formally, the coefficients A_n may be obtained by a contour integration around a small circle centered at the origin:

$$A_n = \frac{1}{2\pi i} \oint \frac{Y}{X^{n+1}} \, dX = \frac{1}{2\pi i} \oint \frac{Y}{X^{n+1}} \frac{dX}{dY} \, dY = \frac{1}{2\pi i} \oint \frac{Y}{f(Y)^{n+1}} f'(Y) \, dY. \qquad (5.42)$$

The crucial step is to transform from integration over X to integration over Y. The transformed contour is also a small circle about the origin, since Y and X are proportional to each other near the origin.

Let's apply this inversion formula to Eq. (5.41). Writing $Y = \mathcal{E}t$ and $X = te^{y-t}$, Eq. (5.41) becomes $f(Y) = Y e^{-Y} = X$. From Eq. (5.42) and using $dX/dY = (1 - Y) e^{-Y}$, we have

$$A_n = \frac{1}{2\pi i} \oint \frac{Y}{(Y e^{-Y})^{n+1}} (1 - Y) e^{-Y} \, dY = \frac{1}{2\pi i} \oint \frac{1 - Y}{Y^n} e^{nY} \, dY.$$

To find the residue we simply expand the exponential in a power series and then read off the coefficient of $1/Y$ in the integral. Thus

$$A_n = \frac{1}{2\pi i} \oint \sum_{k \geq 0} \frac{n^k}{k!} \left(Y^{k-n} - Y^{k+1-n} \right) dY = \frac{n^{n-1}}{(n-1)!} - \frac{n^{n-2}}{(n-2)!} = \frac{n^{n-1}}{n!},$$

so that the series representation of the inverse function $Y(X)$ is

$$Y = \sum_{n \geq 1} \frac{n^{n-1}}{n!} X^n. \qquad (5.43)$$

The generating function itself is obtained by the Lagrange inversion formula (see the highlight above). Identifying $Y = \mathcal{E}t$ and $X = te^{y-t}$ in Eq. (5.43), the result is

$$\mathcal{E}t = \sum_{k \geq 1} \frac{k^{k-1}}{k!} t^k e^{-kt} e^{yk} .$$

Since the density c_k equals the kth term in the series expansion of $\mathcal{E}$ divided by k, we obtain the remarkably simple result

$$c_k(t) = \frac{k^{k-2}}{k!} t^{k-1} e^{-kt} . \qquad (5.44)$$

Using Stirling's approximation, the asymptotic behavior of this distribution is

$$c_k(t) \simeq \frac{k^{k-2}}{\sqrt{2\pi k}} \left(\frac{e}{k}\right)^k t^{k-1} e^{-kt} \xrightarrow[t \uparrow 1]{} \frac{e^{-k(1-t)^2/2}}{\sqrt{2\pi} \, k^{5/2}} , \tag{5.45}$$

where we approximate $e^{-k(t-\ln t - 1)}$ by $e^{-k(1-t)^2/2}$ for $t \to 1$ from below. For $t \neq 1$, the mass distribution c_k decreases exponentially with k. At the gelation time $t = t_g = 1$, the mass distribution has a power-law tail, which signals the singularity where a gel first appears. For t close to t_g, (5.45) gives the scaling form for the mass distribution

$$c_k(t) \simeq s^{-5/2} \Phi(k/s) \qquad \text{with} \quad \Phi(\xi) = \frac{1}{\sqrt{2\pi}} \frac{e^{-\xi/2}}{\xi^{5/2}}, \tag{5.46}$$

in which the characteristic mass is $s = (1-t)^{-2}$.

The generating function allows us to determine the fraction of the total mass that belongs to the gel. Using the implicit solution (5.41) for the generating function we find that $g = 1 - M_1 = 1 - \mathcal{E}(y = 0, t)$ satisfies

$$g = 1 - e^{-gt} . \tag{5.47}$$

This equation always admits a trivial solution $g = 0$. For $t > 1$, however, there is an additional non-trivial solution in which the gel has a non-zero mass; this solution gives the correct value of g for $t > t_g = 1$. While Eq. (5.47) is not explicitly soluble, the limiting behaviors of Eq. (5.47) can be obtained perturbatively. Just past the gelation time, we write $t = 1 + \delta$ and expand (5.47) for small δ, while for $t \to \infty$, we write $g = 1 - \epsilon$ and expand for small ϵ. These considerations give the limiting forms:

$$g = \begin{cases} 0, & \text{for} \quad t < 1, \\ 2(t-1) - 8(t-1)^2/3 + \cdots, & \text{for} \quad t \downarrow 1, \\ 1 - e^{-t} - te^{-2t} + \cdots, & \text{for} \quad t \to \infty. \end{cases} \tag{5.48}$$

Combining (5.38)–(5.39) with the results of (5.48) for g, the cluster density decays as

$$M_0 = N = \begin{cases} 1 - t/2, & \text{for} \quad t \leq 1, \\ 1 - t/2 + 2(t-1)^3/3 + \cdots, & \text{for} \quad t \downarrow 1, \\ e^{-t} + (t/2) e^{-2t} + \cdots, & \text{for} \quad t \to \infty. \end{cases} \tag{5.49}$$

The cluster density, as well as its first two derivatives, remain continuous across the gel point, while the third derivative exhibits a jump.

Finally, we can obtain complete results about moments of the cluster size distribution for *finite* clusters, which will be finite except at the gel point. For example, to determine M_2

we could use the rate equation

$$\frac{dM_2}{dt} = \frac{1}{2} \sum_{i \geq 1} \sum_{j \geq 1} (i+j)^2 \, ij \, c_i \, c_j - \sum_{k \geq 1} k^3 \, c_k$$

$$= \sum_{i \geq 1} \sum_{j \geq 1} i^2 \, c_i j^2 \, c_j + \sum_{i \geq 1} i^3 \, c_i \sum_{j \geq 1} jc_j - \sum_{k \geq 1} k^3 \, c_k$$

$$= M_2^2 - gM_3 \, .$$

Before the gel point we recover the already known solution $M_2(t) = (1-t)^{-1}$. For $t > t_g$, the above equation involves g and M_3 which are not known. Consequently, this approach fails to give an explicit expression for M_2 or for the moments M_n with $n \geq 2$ in the post-gel regime. Nevertheless, we can express the moments for the population of finite clusters directly in terms of g through derivatives of the generating function $\mathcal{E}$:

$$M_n = \left. \frac{\partial^{n-1} \mathcal{E}}{\partial y^{n-1}} \right|_{y=0} \, .$$

For the second moment, we take the logarithm of Eq. (5.41), differentiate with respect to y, and set $y = 0$ to give

$$M_2(t) = \left[\frac{1}{\mathcal{E}(y=0, t)} - t \right]^{-1}, \tag{5.50}$$

with $\mathcal{E}(y=0, t) = 1$ in the sol phase and $\mathcal{E}(y=0, t) = 1 - g = e^{-gt}$ (Eq. (5.47)) in the gel phase. Therefore,

$$M_2(t) = \begin{cases} (1-t)^{-1}, & \text{for} \quad t < 1, \\ (e^{gt} - t)^{-1}, & \text{for} \quad t > 1. \end{cases} \tag{5.51}$$

For $t \to t_g$ from below, the second moment grows rapidly with time, while for large t, $M_2 \to 0$ as finite clusters are progressively engulfed by the gel (Fig. 5.4). Higher moments exhibit similar qualitative behavior (problem 5.3).

Other idealized kernels

The most general form of the polymerization kernel in (5.33) may be written as

$$K_{ij} = A + B(i+j) + Cij, \tag{5.52}$$

i.e. an arbitrary linear combination of the constant, sum, and product kernels. This kernel represents most exactly soluble aggregation models.[7] The solvability of the general

---

[7] The remaining exactly solved models include the exponential kernel, $K_{ij} = 2 - q^i - q^j$ with $0 < q < 1$, that interpolates between the constant kernel, $K_{ij} = 2$ when $q = 0$, and the sum kernel, $K_{ij} \simeq (1-q)(i+j)$, when $q \to 1 - 0$. The other exactly solved model is the parity-dependent kernel where K_{ij} takes three distinct values depending on whether i, j are both even, both odd, or one index is even and the other is odd.
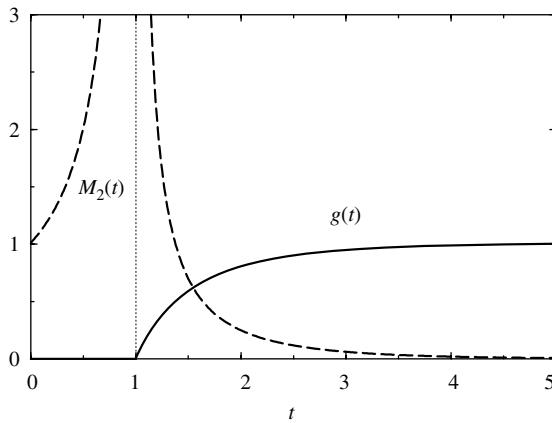
Fig. 5.4. Time dependence of the mass of gel phase (solid) and the second moment of the sol phase (dashed).

polymerization kernel cannot be deduced by "superposition" however, because the master equations are nonlinear. Nevertheless, since the constant, sum, and product kernels are separately solvable, and because the sum and product kernels are intimately related (see problem 5.4), it is plausible that the general kernel (5.52) is tractable. This solution confirms natural expectations: for $C > 0$, the product term dominates and the system undergoes gelation; if $C = 0$ and $B > 0$, the sum term dominates and the typical mass grows exponentially with time, as in the sum-kernel solution.

Kernels with more rapid aggregation, such as

$$K_{ij} = (i+j)^2 \quad \text{or} \quad K_{ij} = (ij)^2 \,, \tag{5.53}$$

are much more challenging. Formally, these kernels appear amenable to generating function techniques, in which the master equations can be recast as a single partial differential equation for the generating function. However, this equation is nonlinear and higher than first order (in contrast to the first-order equation (5.40) for the product kernel). To make progress, it is natural to use the moment method to extract basic information, such as the typical mass, or to detect the emergence of the gel. Unfortunately, the moment method is not useful for kernels such as (5.53). To see the cause of the trouble, consider the kernel $K_{ij} = (i+j)^2$, for which the equations for the first few moments are:

$$\begin{aligned}
\dot{M}_2 &= 2M_1 M_3 + 2M_2^2, \\
\dot{M}_3 &= 3M_1 M_4 + 9M_2 M_3, \\
\dot{M}_4 &= 4M_1 M_5 + 14M_2 M_4 + 10M_3^2, \\
\dot{M}_5 &= 5M_1 M_6 + 20M_2 M_5 + 35M_3 M_4 \,.
\end{aligned} \tag{5.54}$$

Equations (5.54) are *hierarchical* – the equation for each $\dot{M}_n$ contains a *higher* moment M_{n+1} – and are generally insoluble. In previous examples, the moment equations were

recurrent, and hence solvable one by one. Moreover, this hierarchical structure leads to the pathology, for specific examples, such as $K_{ij} = (i+j)^2$, that a gel appears *instantaneously* $-t_g = 0^+$! This strange behavior also occurs for models with asymptotically homogeneous kernels of the form

$$K_{1,j} = K_{j,1} \sim j^{\nu} \quad \text{when} \quad j \to \infty \tag{5.55}$$

with ν strictly greater than 1. However, the applicability of such models to real aggregation processes is questionable[8] because the number of active sites on a cluster should not increase faster than its size. This implies that the ratio $K(1,j)/j$ is bounded as $j \to \infty$, leading to the bound $\nu \leq 1$. The proof that $t_g = 0^+$ in aggregation with the kernel (5.55) and with $\nu > 1$ is cumbersome. We shall show in Section 5.6 that instantaneous gelation arises in mass exchange processes; this analysis is simpler than that for conventional aggregation.

5.4 Scaling

Scaling is based the observation that the typical cluster mass grows systematically with time. Thus a change in time scale corresponds to a change in mass scale, so that the mass distribution becomes invariant when the mass is scaled appropriately with time. This invariance is embodied by the *scaling ansatz*

$$c(x, t) = \frac{1}{s^2} f\left(\frac{x}{s}\right). \tag{5.56}$$

Here $s = s(t)$ is the typical cluster mass at time t, x/s is the scaled mass, and $f(x/s)$ is the scaling function. Thus the fundamental variables are not the mass and time separately, but rather the scaled mass x/s and the time. The prefactor s^{-2} in the scaling function is needed to ensure mass conservation: $\int x\, c(x, t)\, dx = 1$. In scaled units this conservation statement reduces to the manifestly time-independent relation $\int u f(u)\, du = 1$.

  Scaling plays a major role to help understand numerous non-equilibrium phenomena. There are several reasons for the utility of the scaling approach:

- Scaling simplifies and condenses the description of a non-equilibrium system.
- Scaling provides the simplest route to the asymptotic solution when exact solutions are difficult or impossible to obtain because scaling separates a two-variable master equation into single-variable equations that can be analyzed individually.
- A scaling solution is independent of the initial conditions[9] and thus focuses on the most relevant asymptotic behavior.

---

[8] Kernels with $\nu > 1$ have been proposed, for example, for gravitationally attracting particles. Such kernels can, in principle, be used for finite systems. In the thermodynamic limit, however, aggregation with such kernels is ill-defined as the evolution stops instantaneously.

[9] Apart from a trivial dependence on the total mass (see e.g. Eq. (5.24)). If, however, the initial mass distribution has a sufficiently slowly decaying tail (see Example 5.3 in Section 5.2), then the asymptotic solution depends on the initial conditions in an essential way.

- Finally, scaling provides a robust way to classify solutions in terms of generic features of the reaction kernel. We will also see this approach in action in later chapters.

To apply scaling, we need a "right" definition of the typical mass. From the scaling ansatz, the nth moment of the mass distribution is

$$M_n = \int x^n \, \frac{1}{s^2} f\left(\frac{x}{s}\right) \, dx \sim s^{n-1} \, .$$

Consequently, $1/M_0$, M_2, or even the ratios M_{n+1}/M_n for arbitrary n are reasonable measures of the typical mass s, as long as the mass distribution itself is not too singular. When scaling holds, we can define the typical mass to best suit the particular situation at hand. We also need basic information about the matrix of reaction rates K_{ij} to determine the consequences of scaling. It turns out that only two features of this matrix determine the asymptotic properties of the mass distribution. The first is the homogeneity index, λ, defined by

$$K(ai, aj) \sim a^\lambda K(i, j),$$

that gives the overall mass dependence of the reaction rate; here we write arguments rather than subscripts for K because i and j will be taken as continuous. The second is the index ν, defined in (5.55), that characterizes the relative importance of reactions between clusters of disparate masses. As examples, the constant kernel has indices $(\lambda, \nu) = (0, 0)$, the product kernel has $(\lambda, \nu) = (2, 1)$, and the "Brownian" kernel (5.5), which gives the reaction rate for spherical diffusing aggregates, has indices $(\lambda, \nu) = (0, 1/3)$.

The role of the indices λ and ν on the reaction kinetics may be appreciated by considering the following pictorial representation of the reaction matrix.

$$K_{ij} = \begin{pmatrix} \begin{array}{c|c} SS & SL \\ \hline LS & LL \end{array} \\ \end{pmatrix}$$

The meta-entries SS, $SL(= LS)$, and LL denote the reaction rates of small clusters with other small clusters, large–small interactions, and large–large interactions, respectively. The exactly soluble examples discussed in Section 5.2 represent archetypes of three distinct universality classes with the following general behavior:

- *Type I: LL* $\gg$ *LS, SS, corresponding to* $\lambda > \nu$. For this case, the high reactivity of large clusters causes them to quickly disappear, while small clusters tend to persist. Consequently, the cluster mass distribution decays monotonically with mass. The product kernel typifies this class of reaction rates.
- *Type II: all three reactions are of the same order*. This marginal class contains the constant kernel $K_{ij} = 1$. The asymptotic behavior of this class is sensitive to details of the reaction rates.
- *Type III: LS* $\gg$ *LL, SS, or* $\lambda < \nu$. As the reaction develops, small clusters are quickly depleted because of the dominance of large–small interactions. Thus the system has a dearth of small clusters, leading to a peaked mass distribution.

Let's now apply scaling to determine basic features of the cluster mass distribution. In the continuum limit, the master equations for aggregation are[10]

$$\frac{\partial c}{\partial t} = \frac{1}{2} \int_0^x dy\, K(y, x-y)\, c(y,t)\, c(x-y,t) - \int_0^\infty dy\, K(x,y)\, c(x,t)\, c(y,t). \quad (5.57)$$

Substituting the scaling form $c(x,t) = s^{-2} f(x/s)$ into Eq. (5.57), and using homogeneity, $K(ax, ay) = a^\lambda K(x,y)$, the left-hand side becomes

$$\frac{\partial c}{\partial t} = -\frac{\dot{s}}{s^3} \left[2f(u) + uf'(u) \right],$$

where $u = x/s$. Similarly, the right-hand side equals $s^{\lambda-3}\, \mathcal{K}(u)$, where

$$\mathcal{K}(u) = \frac{1}{2} \int_0^u dv\, K(v, u-v) f(v) f(u-v) - \int_0^\infty dv\, K(u,v) f(u) f(v), \quad (5.58)$$

and with $v = y/s$. Equating these two sides and rearranging gives

$$\frac{\dot{s}(t)}{s(t)^\lambda} = -\frac{\mathcal{K}(u)}{2f(u) + uf'(u)} \equiv \Lambda. \quad (5.59)$$

Thus, as emphasized previously, the dependences on time and on scaled mass u separate and the two sides of Eq. (5.59) must separately equal a constant – the separation constant Λ.

The time dependence of the typical mass is determined from $\dot{s} = \Lambda s^\lambda$ and gives

$$s(t) \sim \begin{cases} t^{1/(1-\lambda)} \equiv t^z, & \lambda < 1, \\ e^{\Lambda t}, & \lambda = 1, \\ (t_g - t)^{-1}, & 1 < \lambda \leq 2. \end{cases} \quad (5.60)$$

For non-gelling systems, the time dependence of the typical mass is determined *only* by the homogeneity index λ; other features of the reaction rate, such as the secondary index ν, affect only details, such as the overall amplitude in the above asymptotic decay laws. The power-law time dependence in Eq. (5.60) can also be obtained from a simple argument. If $s(t)$ represents the typical cluster mass at time t, the corresponding density of clusters is of the order of $1/s$. Consider now a time interval Δt during which all clusters react. The typical mass therefore increases by $\Delta s \sim s$, while the time needed for this set of events to occur scales as the inverse of an overall reaction rate. In turn, the latter is proportional to the reaction kernel $K(s,s) \sim s^\lambda$ times the concentration $1/s$. Hence $\Delta s/\Delta t \sim s \times (s^\lambda/s)$, which reproduces (5.60).

The dependence of the scaling function f on the scaled mass u is governed by the u-dependent part of (5.59),

$$2f(u) + uf'(u) + \Lambda^{-1}\mathcal{K}(u) = 0, \quad (5.61)$$

---

[10] The lower limit of the integrals in Eq. (5.57) should actually be equal to the smallest initial mass. As long as the cluster mass distribution is not too singular near the origin, we may replace this lower limit by 0.

with $\mathcal{K}(u)$ given by (5.58). This nonlinear integro-differential equation (5.61) is complicated, and a full understanding of the behavior of the scaling function $f(u)$ is still lacking. It is not possible to solve (5.61) for arbitrary kernels, but we can understand various qualitative features by simple means. For example, it has been shown that when the scaled mass is large, $u \gg 1$, the scaling function has an exponential dependence $f(u) \sim e^{-au}$; the detailed properties of the kernel affects only the prefactor of this exponential tail. The behavior of the scaling function when the scaled mass is small is less robust. Various numerical results indicate that $f(u) \sim u^{-\tau}$ for $u \ll 1$. As a corollary, the density of small-mass clusters has the time dependence

$$c_k \sim \frac{1}{s^2} \left(\frac{k}{s} \right)^{-\tau} \sim k^{-\tau} t^{-(2-\tau)z} . \tag{5.62}$$

A heuristic approach to determine τ is to *assume* that the time dependence of the monomers is representative of the scaling function in the $u \to 0$ limit. The evolution equation for the monomer density is

$$\frac{dc_1}{dt} = -c_1 \sum_{j \geq 1} K_{1j} c_j. \tag{5.63}$$

Since $K_{1j} \sim j^{\nu}$ and $c_j \sim j^{-\tau}$, the sum on the right-hand side of (5.63) converges if $\nu - \tau + 1 < 0$. In this case the first term provides a good estimate for the sum. Thus we write $\dot{c}_1 \sim -K_{11} c_1^2$, leading to $c_1 \sim 1/t$. Matching this time dependence with Eq. (5.62), we deduce the exponent relation $\tau = 1 + \lambda$. This power-law tail applies for $\nu - \tau + 1 = \nu - \lambda < 0$, that is, for Type I kernels. This time dependence for small-mass clusters is one well-established feature of the tail of the cluster size distribution.

5.5 Aggregation with input

Many physical realizations of aggregation do not occur in a closed system, but instead a steady input helps drive the reaction. Examples of aggregation with input are diverse, ranging from chemical processing in continuously stirred tank reactors to the distribution of star masses in the galaxy. Here we study the interplay between input and aggregation in the simple case of a constant monomer input that begins at $t = 0$. Because the asymptotic behavior is independent of initial conditions, we treat an initially empty system: $c_k(0) = 0$.

Constant kernel

The mass distribution now evolves according to

$$\frac{dc_k}{dt} = \sum_{i+j=k} c_i c_j - 2c_k N + \delta_{k,1} . \tag{5.64}$$

The total density satisfies $\dot{N} = -N^2 + 1$ whose solution, for an initially empty system, is

$$N(t) = \tanh t. \tag{5.65}$$

Hence the density of clusters initially grows linearly with time but eventually saturates to 1.

The densities of clusters of a given mass can in principle be found by solving the master equations one by one. However, again the generating function approach is a more potent tool. Introducing the generating function $\mathcal{C}(z,t) = \sum_{k\geq1} c_k(t)z^k$, we recast Eqs (5.64) into the differential equation (compare with Eq. (5.21))

$$\dot{\mathcal{C}}(z,t) = \mathcal{C}(z,t)^2 - 2\mathcal{C}(z,t)N(t) + z. \tag{5.66}$$

As in irreversible aggregation, it is convenient to work with the quantity $\mathcal{C} - N$ that satisfies a closed differential equation $d(\mathcal{C}-N)/dt = (\mathcal{C}-N)^2 + (z-1)$. Solving this equation gives

$$\mathcal{C}(z,t) = \tanh t - \sqrt{1-z}\,\tanh\left(t\sqrt{1-z}\right) \tag{5.67}$$

and therefore in the limit $t \to \infty$ we have

$$\mathcal{C}(z,t=\infty) = 1 - \sqrt{1-z}.$$

The generating function at infinite time can be inverted by expanding $\sqrt{1-z}$ in a power series in z and the tail of the steady-state mass distribution is[11]

$$c_k \simeq \frac{1}{\sqrt{4\pi}}\frac{1}{k^{3/2}}, \qquad k \gg 1. \tag{5.68}$$

Fun with gamma functions

The Euler gamma function provides a convenient and powerful way to invert the generating function for aggregation with input. Consider the function $\sqrt{1-z}$ that appears in the generating function at infinite time. The Taylor expansion of this function is

$$\sqrt{1-z} = 1 + \tfrac{1}{2}(-z) + \tfrac{1}{2}\left(-\tfrac{1}{2}\right)\frac{(-z)^2}{2!} + \tfrac{1}{2}\left(-\tfrac{1}{2}\right)\left(-\tfrac{3}{2}\right)\frac{(-z)^3}{3!}$$
$$+ \tfrac{1}{2}\left(-\tfrac{1}{2}\right)\left(-\tfrac{3}{2}\right)\left(-\tfrac{5}{2}\right)\frac{(-z)^4}{4!} + \cdots.$$

This form is inconvenient for determining the behavior of an arbitrary term in this series. However, we can eliminate the long strings of products by using a basic identity of the gamma function,

$$a(a+1)\ldots(a+k-1) = \frac{\Gamma(a+k)}{\Gamma(a)}, \tag{5.69}$$

---

[11] In this section we write c_k instead of $c_k(t=\infty)$; whenever we treat a non-steady mass distribution, we write $c_k(t)$.

to re-express the above power series in the compact form:

$$\sqrt{1-z} = 1 - \sum_{k \geq 1} \frac{\Gamma\left(k - \frac{1}{2}\right)}{2\Gamma\left(\frac{1}{2}\right)} \frac{z^k}{\Gamma(k+1)} \, .$$

Finally, we use $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ to obtain

$$c_k = \frac{1}{\sqrt{4\pi}} \frac{\Gamma\left(k - \frac{1}{2}\right)}{\Gamma(k+1)} \, . \tag{5.70}$$

For the asymptotic behavior, we may use the handy asymptotic relation for $k \gg 1$,

$$\frac{\Gamma(k+a)}{\Gamma(k+b)} \simeq k^{a-b} \, . \tag{5.71}$$

Using Eqs (5.70) and (5.71), we obtain the asymptotic steady-state mass distribution (5.68). See the highlight on page 390 for another application of gamma functions.

At finite time, mass conservation requires that $\sum k\, c_k(t) = t$, so that the stationary power-law tail of the mass distribution must eventually be cut off. We may determine the crossover between the stationary regime for $k < k_*$ and the regime $k > k_*$ where clusters have not yet been created by requiring that the total mass in the system is proportional to the elapsed time:

$$t = \sum_{k \geq 1} k c_k(t) \sim \sum_{k=1}^{k_*} k c_k \sim \sum_{k=1}^{k_*} k^{-1/2} \sim (k_*)^{1/2} \, , \tag{5.72}$$

Thus $k_* \sim t^2$. Consequently, the steady-state power-law distribution (5.70) extends only up to this cutoff point.

This qualitative picture can be sharpened by an exact analysis. To extract the cluster densities from the generating function (5.67), we substitute the series representation

$$\pi \tanh(\pi x) = \sum_{n=-\infty}^{\infty} \frac{x}{x^2 + \left(n + \frac{1}{2}\right)^2}$$

into (5.67) and expand in powers of z to yield

$$c_k(t) = \frac{1}{t^3} \sum_{n=-\infty}^{\infty} \left(n + \frac{1}{2}\right)^2 \pi^2 \left[1 + \left(n + \frac{1}{2}\right)^2 \frac{\pi^2}{t^2}\right]^{-k-1} \, . \tag{5.73}$$

In the long-time limit, we replace the sum on the right-hand side of (5.73) by an integral over the variable $x = \left(\frac{1}{2} + n\right) \pi / t$. When $k \ll t^2$, we obtain

$$c_k \simeq \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x^2 \, dx}{(1 + x^2)^{k+1}} \; .$$

Evaluating the integral we recover (5.70).

On the other hand, when k and t both diverge such that $\kappa = k/t^2$ remains finite, (5.73) simplifies to

$$c_k(t) = \frac{1}{t^3} \sum_{n=-\infty}^{\infty} \left(n + \frac{1}{2}\right)^2 \pi^2 \exp\left[-\left(n + \frac{1}{2}\right)^2 \pi^2 \kappa\right] \; . \qquad (5.74)$$

For mass $k \gg k_*$, we may keep only the first term in (5.74) to give the leading asymptotic behavior of the mass density,
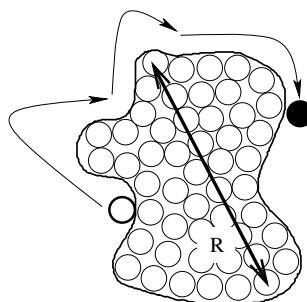
$$c_k(t) = \frac{\pi^2}{4 \, t^3} \, e^{-\pi^2 \kappa / 4} \; , \qquad (5.75)$$

so that large-mass clusters with $k \gg t^2$ are exponentially rare.

Generalized sum kernel

A natural realization of aggregation with input occurs in epitaxial surface growth. Here, atoms impinge onto a clean surface at a constant rate, irreversibly adsorb, and in a suitable temperature range can subsequently diffuse along the surface. These mobile adatoms can merge to form dimers, trimers, etc., so that islands of all sizes can diffuse on the surface and continue to aggregate. Island growth is therefore driven by the combined effects of irreversible aggregation, island diffusion, and steady monomer input.

The mechanism underlying island diffusion is that adatoms on the edge of an island can hop to neighboring sites on the periphery (Fig. 5.5). Consider a large island of linear size

Fig. 5.5.   Schematic illustration of effective island diffusion by the motion of an adatom at the periphery of an island.

R. In a time interval $\Delta t \sim R^2$, an edge adatom can explore the entire island boundary.[12] In this time interval Δt, each periphery adatom moves by a distance R from its original position. The center-of-mass displacement of the island due to the motion of a single adatom is $\delta x \sim R/R^2 \sim R^{-1}$. If each periphery adatom diffuses independently, the total center-of-mass displacement Δx will be the sum of R independent identically distributed random variables, so that $\Delta x \sim \sqrt{R(\delta x)^2} \sim R^{-1/2}$. Thus the effective diffusion coefficient of an island of linear dimension R scales as $D_R \sim (\Delta x)^2/\Delta t \sim R^{-3}$. Since the mass k of an island of linear size R is proportional to its area, the effective diffusion coefficient of the island has the mass dependence

$$D_k \sim k^{-3/2} \, . \tag{5.76}$$

With Eq. (5.76), the Smoluchowski formula for the aggregation rate of an i-mer and j-mer on the surface, $K_{ij} = (D_i + D_j)(R_i + R_j)^{d-2}$, reduces to $K_{ij} \sim D_i + D_j$. Here R_j is the radius of a j-mer and D_j is its diffusion coefficient.[13] We now write $D_j \sim j^{-a}$, where a is the mobility exponent that equals 3/2 for periphery diffusion, but may be different for other surface relaxation mechanisms. In the physical case of a two-dimensional surface, we obtain the *generalized sum kernel*

$$K_{ij} = i^{-a} + j^{-a} \, . \tag{5.77}$$

Let's determine the steady state for the generalized sum kernel with monomer input. The master equations are

$$\frac{dc_k}{dt} = \frac{1}{2} \sum_{i+j=k} (i^{-a} + j^{-a}) \, c_i c_j - c_k \sum_{j\geq 1} (i^{-a} + k^{-a}) c_j + \delta_{k,1} \, . \tag{5.78}$$

Introducing the generating functions

$$\mathcal{A}(z) = \sum_{k\geq 1} k^{-a} c_k z^k \, , \quad \mathcal{C}(z) = \sum_{k\geq 1} c_k z^k \, , \tag{5.79}$$

and the shorthand $A = \mathcal{A}(z = 1)$, we reduce Eq. (5.78) to

$$\mathcal{A}(z)\mathcal{C}(z) - \mathcal{A}(z)N - \mathcal{C}(z)A + z = 0 \, . \tag{5.80}$$

Since the mass distribution decays algebraically when $a = 0$, it is natural to assume that this decay holds when $a \neq 0$; that is, $c_k \to C k^{-\tau}$ when $k \gg 1$. This hypothesis implies that the singular behavior of the generating functions are, as $z \uparrow 1$ (see Eq. (5.26) in the highlight on page 142),

$$\mathcal{A}(z) = A + C\Gamma(1 - \tau - a) \, (1 - z)^{\tau + a - 1} + \cdots$$
$$\mathcal{C}(z) = N + C\Gamma(1 - \tau) \, (1 - z)^{\tau - 1} + \cdots \, . \tag{5.81}$$

---

[12] We tacitly assume that periphery diffusion smooths the boundary so that the perimeter of the island scales as R and that the adatom stays close to the island boundary.

[13] For diffusion on a two-dimensional surface, the term $(R_i + R_j)^{d-2}$ in the Smoluchowski formula should be replaced by the slowly varying factor $1/\ln(R_i + R_j)$ that we ignore here.

To extract the decay exponent τ and the amplitude C, we substitute the above expansions for $\mathcal{A}$ and $\mathcal{C}$ into (5.80) and require consistency in the $z \to 1$ limit. Matching the constant terms yields $AN = 1$. Matching then the first correction terms in $(1 - z)$ and using the reflection formula for the gamma function, $\Gamma(z)\Gamma(1 - z) = \pi / \sin(\pi z)$, we find

$$\tau = \frac{3 - a}{2}, \qquad C = \sqrt{\frac{1 - a^2}{4\pi}} \, \cos\left(\frac{\pi a}{2}\right). \tag{5.82}$$

Having found the decay exponent of the steady-state k-mer concentrations, we now estimate the crossover time by the same reasoning that led to (5.72):

$$t = \sum_{k \geq 1} k c_k(t) \sim \sum_{k=1}^{k_*} k c_k \sim \sum_{k=1}^{k_*} k^{1-\tau} \sim k_*^{2-\tau} = k_*^{(1+a)/2} \, .$$

Thus we infer that the characteristic mass is

$$k_* \sim t^{2/(1+a)} \, . \tag{5.83}$$

Our analysis relied on the assumption that the system reaches a stationary state. Specifically, we tacitly assumed that the sums $N = \sum c_k$ and $A = \sum k^{-a} c_k$ converge, which means that the decay exponent should simultaneously satisfy $\tau > 1$ and $\tau > 1 - a$. These inequalities, in conjunction with $\tau = (3 - a)/2$, impose bounds on the mobility exponent,

$$-1 < a < 1 \, , \tag{5.84}$$

for the steady state to exist. Paradoxically, the periphery diffusion mechanism that motivated the generalized sum kernel does not reach a steady state but instead evolves continuously.

Island growth model

The extreme limit of $a \to \infty$, where only monomers are mobile, while larger aggregates (islands) are immobile (Fig. 5.6), is especially interesting. In this *island growth model*,
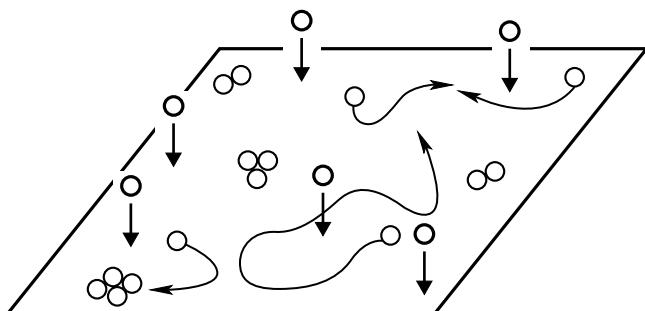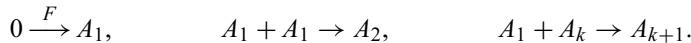
Fig. 5.6. Illustration of the elemental processes in the island growth model including adsorption at rate *F* (vertical arrows), collision of two mobile monomers to make an immobile dimer (upper right), and the incorporation of a mobile monomer into an immobile cluster (lower left).

clusters of mass 2 or greater can grow only by the *addition* of mobile adatoms to their boundaries; there is no aggregation *per se*. The elemental steps of island growth are

$$0 \xrightarrow{F} A_1, \qquad A_1 + A_1 \to A_2, \qquad A_1 + A_k \to A_{k+1}.$$

The first process accounts for the deposition of monomers onto the surface with rate F, while the latter two account for cluster growth as additional monomers are incorporated. When only monomers are mobile, the reaction kernel is $K_{ij} \sim D_i + D_j = D(\delta_{i,1} + \delta_{j,1})$, so that the evolution of the system is governed by the equations

$$\frac{dc_1}{dt} = -c_1^2 - c_1 \sum_{k=1}^{\infty} c_k + F,$$

$$\frac{dc_k}{dt} = c_1(c_{k-1} - c_k), \quad k \geq 2. \tag{5.85}$$

Here we absorb the diffusivity D into the time variable so that the parameter F becomes the ratio of the deposition rate (measured in number of adatoms per site per unit time) to the diffusivity D. In experimental applications, F is usually small and can be varied over a wide range.

The assumptions underlying the master equation description of the island growth model include the following:

- Islands are dilute so that they cannot merge directly. This limit corresponds to the submonolayer regime defined by $Ft \ll 1$.
- Islands are point-like. This is a mild assumption because the reaction rate depends only logarithmically on the island radius in two dimensions.
- The mean-field approximation holds. This assumption cannot really be justified because immobile islands can be viewed as randomly distributed "traps," for which a mean-field description fails (see Section 13.2).
- There are freely diffusing adatoms and stable, immobile islands. This assumption is certainly questionable as islands typically diffuse by the detachment and reattachment of monomers on the island periphery.

In spite of these idealizations, the island growth model nicely illustrates the use of asymptotic analysis to solve the master equations, as well as the power and the limitations of a scaling analysis.

To solve Eqs (5.85), let us first sum them to obtain the rate equation for the total density N:

$$\frac{dN}{dt} = F - c_1 N. \tag{5.86}$$

It is also helpful to explicitly partition clusters into monomers and immobile islands – those of mass 2 or greater. Let I be the density of immobile islands, $I = \sum_{k \geq 2} c_k$. Then monomers

and islands satisfy the coupled equations

$$
\begin{aligned}
\frac{dc_1}{dt} &= -2c_1^2 - c_1 I + F, \\
\frac{dI}{dt} &= c_1^2,
\end{aligned}
\tag{5.87}
$$

whose asymptotic solution is given by (see problem 5.11)

$$
I(t) \simeq F^{2/3} (3t)^{1/3}, \qquad c_1(t) \simeq F^{1/3}(3t)^{-1/3}.
\tag{5.88}
$$

These growth laws cannot hold *ad infinitum*, however, because eventually the coverage increases beyond submonolayer. The end of the submonolayer regime is defined by $t_{\max} \sim F^{-1}$, which corresponds to $I_{\max} \sim F^{1/3}$.

To solve for the island densities in the submonolayer regime, we introduce the auxiliary time variable $\tau = \int_0^t c_1(t') \, dt'$ to recast Eqs (5.85) for immobile islands ($k \geq 2$) to those of the Poisson process,

$$
\frac{dc_k}{d\tau} = c_{k-1} - c_k.
\tag{5.89}
$$

We now treat k as continuous and approximate the difference by the derivative to obtain the linear wave equation

$$
\left(\frac{\partial}{\partial \tau} + \frac{\partial}{\partial k} \right) c_k(\tau) = 0.
\tag{5.90}
$$

The general solution is $c_k(\tau) = f(\tau - k)$, where f is an arbitrary function that is determined by matching the solution to c_1, which plays a role of a boundary condition. Since $c_1(\tau) = f(\tau - 1)$, the solution simply is $c_{k+1}(\tau) = c_1(\tau - k)$. Using the definition of τ and Eq. (5.88) for c_1, we have

$$
\tau \simeq \frac{1}{2} F^{1/3} (3t)^{2/3}, \qquad c_1(\tau) \simeq F^{1/2}(2\tau)^{-1/2},
\tag{5.91}
$$

which then give

$$
c_k(\tau) \simeq \frac{F^{1/2}}{\sqrt{2(\tau - k)}}.
\tag{5.92}
$$

Thus the island size distribution is described by a propagating wave solution.

The solution of Eq. (5.92) does not provide an accurate description when k is small because the replacement of $c_k - c_{k-1}$ by a derivative is unjustified. The wave solution (5.92) is also inapplicable when $k \geq \tau$. It is possible to resolve the latter problem by a more sophisticated continuum approximation that keeps two terms in the Taylor expansion,

$$
c_{k-1} - c_k \simeq -\frac{\partial c}{\partial k} + \frac{1}{2} \frac{\partial^2 c}{\partial k^2}.
$$

Then instead of the wave equation (5.90), one obtains the convection–diffusion equation

$$\frac{\partial c}{\partial \tau} + \frac{\partial c}{\partial k} = \frac{1}{2} \frac{\partial^2 c}{\partial k^2},\tag{5.93}$$

which admits a well-posed solution for all k. This solution is very close to the wave solution when $k < \tau$; the singularity of the wave solution (5.92) is replaced by a finite-width peak at $k = \tau$ (see problem 5.12).
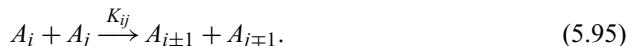
5.6 Exchange-driven growth

Our last topic is the *mass exchange* process, in which the interaction between two clusters is

$$A_i + A_j \xrightarrow{K_{ij;mn}} A_m + A_n.\tag{5.94}$$
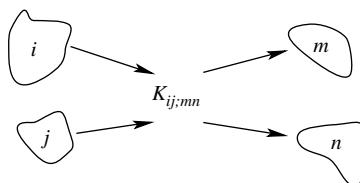
We consider only mass-conserving events in which $i+j = m+n$. Realizations of exchange processes include droplet growth by evaporation and recondensation, and the evolution of wealth by the exchange of capital between economically interacting individuals. As illustrated in Fig. 5.7, mass exchange is a hybrid of scattering and aggregation because two incoming clusters are transformed into two outgoing clusters, while if one of m or n is small, the process resembles aggregation because one of the outgoing clusters has a mass that is nearly equal to the sum of the incoming masses.

As a simple example, suppose that a unit mass is transferred in each exchange event. Then the interaction reduces to

$$A_i + A_j \xrightarrow{K_{ij}} A_{i\pm1} + A_{j\mp1}.\tag{5.95}$$

We further assume that a cluster is equally likely to gain or to lose mass in an interaction so that the reaction matrix is symmetric: $K_{i,j} = K_{j,i}$. A cluster disappears when its mass reaches zero. Thus the number of clusters decreases by 1 with probability 1/2 when one of the reacting clusters is a monomer; if both incoming clusters are monomers, the number of clusters necessarily decreases by 1.

Part of the appeal of exchange processes is that they are readily solvable. This solubility is surprising at first sight, since the moments of the mass distribution (which are so helpful

Illustration of mass exchange, where the interaction of clusters of masses *i* and *j* leads to the creation of clusters of masses *m* and *n*.

in ordinary aggregation) do not satisfy closed equations.[14] However, the master equations for exchange kernels can often be linearized and thereby solved. Consider, for example, the product kernel $K(i,j) = (ij)^\lambda$. The evolution of c_k for this exchange process is governed by

$$\frac{dc_k}{dt} = M_\lambda \left[(k+1)^\lambda c_{k+1} + (k-1)^\lambda c_{k-1} - 2k^\lambda c_k \right], \tag{5.96}$$

where $M_\lambda = \sum_j j^\lambda c_j$ is the λth moment of the mass distribution. These equations apply for all $k \geq 1$ if we impose the constraint $c_0 \equiv 0$. A useful first step to solve Eqs (5.96) is to absorb the factor M_λ into the time variable,

$$\tau = \int_0^t dt' \, M_\lambda(t'), \tag{5.97}$$

thereby linearizing the master equations:

$$\frac{dc_k}{d\tau} = (k+1)^\lambda c_{k+1} + (k-1)^\lambda c_{k-1} - 2k^\lambda c_k. \tag{5.98}$$

We now solve Eqs (5.98) for the illustrative cases of $\lambda = 0, 1, 2$, and show the emergence of instantaneous gelation when $\lambda > 2$.

Example 5.4. *Constant kernel* ($\lambda = 0$). When $K(i,j) = 1$, Eqs (5.98) become

$$\frac{dc_k}{d\tau} = c_{k+1} + c_{k-1} - 2c_k, \tag{5.99}$$

which is just the continuous-time discrete-space diffusion equation. It is helpful to view k as the spatial location of a "particle," so that $c_k(t)$ is the probability that the particle is at k at time t. For the monomer-only initial condition, $c_k(0) = \delta_{k,1}$, the particle starts at $k = 1$ and subsequently diffuses. The constraint $c_0 \equiv 0$ may be interpreted as an absorbing boundary condition.

In the absence of the absorbing boundary condition, the solution for the concentration is $c_k = e^{-2\tau} I_{k-1}(2\tau)$ (see the discussion about discrete-space diffusion in Section 2.1). The shift of the index in the modified Bessel function compared to the solution in Section 2.1 reflects the particle starting at $k = 1$ rather than $k = 0$. We can now incorporate the absorbing boundary condition by putting an image at $k = -1$ and then treating Eq. (5.99) on the entire line,[15] subject to the initial condition $c_k(0) = \delta_{k,1} - \delta_{k,-1}$. Combining the direct and image solutions gives

$$c_k(\tau) = e^{-2\tau} \left[I_{k-1}(2\tau) - I_{k+1}(2\tau) \right], \tag{5.100}$$

with I_ν the modified Bessel function of order ν. From this solution, the total density of clusters, $N(\tau) = \sum_{k \geq 1} c_k(\tau)$, is

$$N(\tau) = e^{-2\tau} \left[I_0(2\tau) + I_1(2\tau) \right] \simeq (\pi\tau)^{-1/2}, \qquad \tau \to \infty.$$

---

[14] The only exception is the first moment, which again is necessarily conserved, $M_1 = 1$.
[15] This standard trick is widely used in electrostatics and it equally applies to the diffusion equation.

This result coincides with the survival probability for diffusion on the infinite half-line. To express this solution in terms of the physical time t, we integrate $t(\tau) = \int_0^\tau d\tau'/N(\tau')$ to obtain, in the long-time limit, $N \simeq [2/(3\pi t)]^{1/3}$, while the mass density is

$$c_k(t) \simeq \frac{k}{3t} \exp\left[-\left(\frac{\pi}{144}\right)^{1/3} \frac{k^2}{t^{2/3}}\right].$$

Example 5.5. *Product kernel* $(\lambda = 1)$. The product kernel $K(i,j) = ij$ turns out to be the simplest solvable exchange model because the first moment M_1 equals 1, so that t and τ in Eq. (5.97) coincide. In this case, Eqs (5.98) become

$$\frac{dc_k}{dt} = (k+1)c_{k+1} + (k-1)c_{k-1} - 2kc_k . \qquad (5.101)$$

Equations (5.101) are readily solved with the exponential ansatz $c_k = Aa^{k-1}$ discussed on page 139. Substituting this ansatz into (5.101), this infinite set of equations reduces to two coupled equations

$$\dot{a} = (1-a)^2, \quad \dot{A} = -2A(1-a),$$

from which $a = t/(1+t)$ and $A = (1+t)^{-2}$. Thus we obtain $c_k = t^{k-1}/(1+t)^{k+1}$. We can obtain this same result with less work by noticing that the condition $\sum k c_k = 1$ leads to $A = (1-a)^2$, and then we only need solve the single differential equation for a.

This mass distribution is precisely the same as that for constant-kernel aggregation (see Eq. (5.13)); this surprising coincidence also holds for exponentially decaying initial data. This coincidence is unique and there is generally no precise mapping between the *nonlinear* master equations (5.6) for aggregation and the *linear* master equations (5.101) for exchange.

Example 5.6. *Quadratic product kernel* $(\lambda = 2)$. The quadratic product kernel $K(i,j) = (ij)^2$ leads to "complete" gelation in which all finite-size cluster densities vanish beyond the gel point. This singularity can be detected by analyzing the rate equations for the moments. In terms of the auxiliary time variable defined in Eq. (5.97), the first few of these equations are[16]

$$\frac{dM_2}{d\tau} = 2M_2, \qquad \frac{dM_3}{d\tau} = 6M_3, \qquad \frac{dM_4}{d\tau} = 12M_4 + 2M_2,$$

and the solutions are combinations of exponentials: $M_2 = e^{2\tau}$, $M_3 = e^{6\tau}$, $M_4 = \frac{6}{5}e^{12\tau} - \frac{1}{5}e^{2\tau}$. Solving for the physical time by $t = \int_0^\tau d\tau' M_2^{-1}(\tau')$ we obtain $t = \frac{1}{2}[1 - e^{-2\tau}]$. Expressing the moments in terms of t gives

$$M_2 = (1-2t)^{-1}, \qquad M_3 = (1-2t)^{-3}, \qquad M_4 = \frac{6}{5}(1-2t)^{-6} - \frac{1}{5}(1-2t)^{-1} .$$
$$(5.102)$$

Generally $M_n \sim (1-2t)^{-n(n-1)/2}$. Therefore gelation occurs at a time $t_g = 1/2$.

---

[16] The total density obeys $dM_0/d\tau = -c_1$ for all kernels, and hence to determine the zeroth moment one must know the monomer density.

Example 5.7. *Instantaneous gelation* ($\lambda > 2$). To establish the unusual (or pathological, depending on taste) phenomenon of instantaneous gelation, let us assume that no gelation occurs and then derive a contradiction. By hypothesis, the pre-gel regime is well defined where the moments exist and evolve according to

$$\frac{dM_n}{d\tau} = 2 \sum_{l=1}^{\lfloor n/2 \rfloor} \binom{n}{2l} M_{n-2l+\lambda} . \tag{5.103}$$

A lower bound for this growth rate is obtained by keeping only the first term in the above sum:

$$\frac{dM_n}{d\tau} \geq n(n-1)M_{n-2+\lambda} . \tag{5.104}$$

This is still a hierarchical set of equations since, for $\lambda > 2$, the index $n-2+\lambda$ of the moment on the right-hand side exceeds the index of the moment on the left-hand side. Fortunately, we can use the *Jensen inequality* (see the highlight, below) to deduce that

$$M_{n-2+\lambda} \geq (M_n)^{1+\mu} , \qquad \mu = \frac{\lambda - 2}{n - 1} . \tag{5.105}$$

Combining (5.104)–(5.105), we obtain $dM_n/d\tau \geq n(n-1)(M_n)^{1+\mu}$. This equation suggests considering the auxiliary functions $\mathcal{M}_n$ that evolve according to

$$\frac{d\mathcal{M}_n}{d\tau} = n(n-1)(\mathcal{M}_n)^{1+\mu} . \tag{5.106}$$

Solving (5.106) subject to the initial condition $\mathcal{M}_n(0) = 1$ yields $\mathcal{M}_n = [1 - n(\lambda - 2)\tau]^{-(n-1)/(\lambda-2)}$. Hence $\mathcal{M}_n \to \infty$ as $\tau \to \tau_n = [n(\lambda - 2)]^{-1}$. Since $M_n \geq \mathcal{M}_n$, the moment M_n diverges at least at τ_n. The series of times τ_n set an upper bound for the gelation time τ_g since all moments should be finite for $\tau < \tau_g$. Since $\tau_n \to 0$ as $n \to \infty$, we conclude that $\tau_g = 0$. Thus for $\lambda > 2$ gelation is instantaneous: $t_g = 0$.

Jensen inequality

Suppose p_j for $j = 1, 2, \ldots$ are non-negative numbers, $p_j \geq 0$, that satisfy the constraint $\sum_j p_j = 1$. Let $\Phi(x)$ be an arbitrary convex function. Then the inequality

$$\sum_{j \geq 1} p_j \Phi(x_j) \geq \Phi\left(\sum_{j \geq 1} p_j x_j \right) \tag{5.107}$$

is called the *Jensen inequality*. This inequality is equivalent to that statement of convexity, i.e. a function is convex if (5.107) is obeyed. Let us substitute $p_j = jc_j$ (the constraint $\sum_j p_j = 1$ then follows from mass conservation $\sum_j jc_j = 1$) and the convex function $\Phi(x) = x^{1+\mu}$ ($\mu > 0$ for $\lambda > 2$) into the Jensen inequality. Choosing $x_j = j^{n-1}$, and using $\sum p_j x_j = \sum j^n c_j = M_n$ as well as $\sum p_j \Phi(x_j) = M_{n-2+\lambda}$, we find that the Jensen inequality reduces to (5.105).

5.7 Notes

Some of the early work on aggregation is reviewed by Chandrasekhar [12], and later in [80, 81] in the context of physical chemistry. Additional applications of aggregation, again primarily from the perspective of physical chemistry, are discussed in [82, 83]. One of the assumptions of the rate equation approach – that the aggregation rate is independent of cluster shape – was validated numerically in [84]. A general discussion of the Stokes–Einstein relation and other aspects of physical hydrodynamics can be found in [85]. Generating functions, which were briefly mentioned in Chapter 2, are used in an essential way in this chapter for solving master equations, in the Lagrange inversion method, and in the analysis of singularities. Three useful books on generating functions include [86–88]. Another more advanced reference is [89] which is particularly useful for singularity analysis. Recent review articles on aggregation, with an emphasis on scaling, can be found in [90, 91].

Early references on aggregation in the presence of a steady source are [92, 93]. The former treats the astrophysical problem of the size distribution of stars due to gravitational accretion of matter and continuous input of new material as a result of supernovae; the latter treats the same phenomena, but in the context of aerosols. Epitaxy models and the connection to aggregation are discussed, for example in [94–97]. The idea underlying mass exchange processes was articulated by Melzak [98]; for the application of this idea to statistical physics modeling of wealth exchange, see [99]. For more information about the subtleties associated with instantaneous gelation, see [100–103]. *The On-Line Encyclopedia of Integer Sequences* mentioned in the problems is [104].

5.8 Problems

5.1 Consider constant-kernel aggregation with the initial mass distribution $c_j(0) = 0$ for $j \geq m + 1$ and with $m > 4$. Deduce the asymptotic behavior of the cluster mass distribution by generalizing the method of Example 5.2 (page 142) as follows:

(a) Although it is not possible to compute the roots of the generating function explicitly, show that the initial generating function $\mathcal{C}(z, t = 0)$ can be written as

$$\mathcal{C}(z_1, 0) = c_1(0)z_1 + c_2(0)z_1^2 + \cdots + c_m(0)z_1^m = 1 + M(0)\epsilon + \mathcal{O}(\epsilon^2),$$

where $c_1(0) + c_2(0) + \cdots + c_m(0) = 1$ is the initial cluster density and $c_1(0) + 2c_2(0) + \cdots + mc_m(0) = M(0)$ is the initial mass density. Here $z_1 = 1 + \epsilon$ is the smallest root of $\mathcal{C}(z_1, 0)$.

(b) Use the above result for $\mathcal{C}(z, t = 0)$ to show that the generating function can be written as

$$\mathcal{C}(z, t) \simeq \frac{1}{t^2 M(0)} \frac{1}{1 - z/z_1},$$

where $z_1 = 1 + 1/[M(0)t]$ is the smallest root in the generating function. Expand this expression in a power series in z and show that the scaling form for the mass distribution is given by Eq. (5.24).

5.2 Solve for the cluster concentrations $c_k(t)$, for constant-kernel aggregation with the initial condition $c_k(0) = 2^{-k}$.

5.3 Show that the third and fourth moments of the cluster mass distribution for product-kernel aggregation with the monomer-only initial condition are given by

$$M_3(t) = \begin{cases} (1-t)^{-3}, & \text{for} \quad t < 1, \\ e^{2gt}(e^{gt}-t)^{-3}, & \text{for} \quad t > 1, \end{cases}$$

and

$$M_4(t) = \begin{cases} (1+2t)(1-t)^{-5}, & \text{for} \quad t < 1, \\ (e^{4gt}+2te^{3gt})(e^{gt}-t)^{-5}, & \text{for} \quad t > 1. \end{cases}$$

More generally, show that near the gel point the moments M_n diverge as

$$M_n \simeq \frac{2^{n-2}\,\Gamma(n-3/2)}{\Gamma(1/2)}\,|1-t|^{-(2n-3)}\,,$$

which follows from Eq. (5.46).

5.4 Consider sum-kernel aggregation in which $K_{ij} = i + j$.

(a) Show that the master equations are

$$\frac{dc_k}{dt} = \frac{k}{2}\sum_{i+j=k} c_i c_j - kc_k(kM_0 + 1)\,,$$

where the mass density equals one: $M_1 = \sum_{k \geq 1} c_k = 1$.

(b) Write the rate equations for the moments. Show that for the monomer-only initial conditions the first few moments are $M_0 = e^{-t}, M_2 = e^{2t}, M_3 = 3\,e^{4t} - 2\,e^{3t}, M_4 = 15\,e^{6t} - 20\,e^{5t} + 6\,e^{4t}$.

(c) Solve the above master equations for the monomer-only initial conditions and obtain the exact solution

$$c_k(t) = \frac{k^{k-1}}{k!}\,(1 - e^{-t})^{k-1}\,e^{-t}\,e^{-k(1-e^{-t})}\,.$$

5.5 Consider the island growth model in the absence of input in which $A_1 + A_k \to A_{k+1}$, where the rate of each reaction equals 1, except for the reaction between monomers, $A_1 + A_1 \to A_2$, whose rate equals 2 (why?). Suppose that the system initially contains only monomers.

(a) Introduce an auxiliary time variable that linearizes the master equations for this reaction.

(b) Determine the island size distribution at infinite time.

(c) Show that the final state is approached exponentially in time.

5.6 Consider the addition process of the previous problem but assume that the reaction kernel is $K_{ij} = i\delta_{j,1} + j\delta_{i,1}$. Show that

$$c_k(t) = \left[(1 - e^{-t})^{k-1} - k^{-1}(1 - e^{-t})^k\right](2 - e^{-t})^{-k}.$$

5.7 Use the scaling approach to determine the mass distribution for constant-kernel aggregation; that is, solve Eq. (5.59) for the scaling function.

5.8 Consider constant-kernel aggregation with steady monomer input.[17]
   (a) Show that the monomer density is given by

   $$c_1 = \frac{1}{2}\left[\frac{t}{\cosh^2 t} + \tanh t\right].$$

   (b) Solve Eqs (5.64) recursively and derive the formal solution (for $k > 1$)

   $$c_k(t) = \frac{1}{\cosh^2 t}\int_0^t dt' \cosh^2 t' \sum_{i+j=k} c_i(t')c_j(t').$$

5.9 Consider sum-kernel aggregation with input.
   (a) Verify that the total density of clusters is given by $N(t) = \int_0^t dt' \, e^{(t'^2 - t^2)/2}$. Show that $N(t)$ initially grows, reaches a maximum, and then decays to zero as $N \simeq t^{-1}$ when $t \gg 1$.
   (b) Show that the density of monomers also initially increases, then decreases, and asymptotically decays as $c_1 \simeq t^{-1}$.
   (c) Show that the density of dimers decays as $c_2 \simeq t^{-3}$.
   (d) Verify that generally $c_k \simeq A_k t^{-(2k-1)}$ in the long-time limit.
   (e) Find the recursion for the amplitudes A_k and show that these amplitudes form the sequence A088716 from *The On-Line Encyclopedia of Integer Sequences*.

5.10 Investigate product-kernel aggregation with input.
   (a) Show that in the pre-gel regime, the second moment is $M_2 = \tan t$. Use this solution to argue that gelation occurs at $t_g = \pi/2$.
   (b) Show that $N = t - t^3/6$ in the pre-gel regime.
   (c) Show that $M_3 = \tan t + \frac{2}{3}\tan^3 t$ in the pre-gel regime.
   (d) Show that the monomer density is $c_1(t) = \int_0^t dt' \, e^{(t'^2 - t^2)/2}$ for all times.

5.11 Starting with the evolution equations (5.87) for the island growth model, determine the time dependence of the island density $I(t)$ and the monomer density $c_1(t)$ by the following steps:
   (a) Show that the island density monotonically increases with time and that $c_1 \to 0$ as $t \to \infty$. *Hint:* Suppose the opposite; namely, $c_1 \to$ const. as $t \to \infty$, and show that this behavior leads to a contradiction.
   (b) Using $c_1 \to 0$ for $t \to \infty$ determine the asymptotic solution for the master equation for c_1. Use this result to derive Eq. (5.88).

---

[17] In this and other problems with input we assume that the system is initially empty, and the input rate is set to one so that mass density is $M = t$.

5.12 Analyze the improved continuum description (5.93) of the island growth model.
   (a) Show that this approach cures the singularity of the wave solution (5.92), namely, the singular behavior is replaced by a finite-width peak.
   (b) Determine the height of the peak, that is, the density $c_k(\tau)$ when $k = \tau$.
   (c) Estimate the width of the peak.

5.13 Investigate the island growth model with unstable dimers, namely each dimer can break into two mobile adatoms with rate λ, while all larger islands are stable.
   (a) Write and solve the master equations for this process, and show that asymptotically $I = F^{3/4}(4t/\lambda)^{1/4}$, $c_1 = F^{1/4}(4t/\lambda)^{-1/4}$, and $c_2 = F^{1/2}(4t\lambda)^{-1/2}$.
   (b) Show that the maximal island density, which is reached at the end of the submonolayer regime, scales with F as $I_{\max} \sim F^{1/2}$.

5.14 Assume that islands of mass $\geq n$ are stable, while lighter islands are unstable.
   (a) Show that $c_1 \sim t^{-1/(n+1)}$.
   (b) Show that $c_n \sim t^{-(n-1)/(n+1)}$.
   (c) Show that $I_{\max} \sim F^{(n-1)/(n+1)}$.

5.15 Consider "greedy" exchange in which the larger of the two interacting clusters always gains a monomer, while the smaller loses a monomer. Symbolically, $(j, k) \to (j + 1, k - 1)$ for $j \geq k$.
   (a) Show that the evolution of the system is described by

$$\dot{c}_k = c_{k-1} \sum_{j=1}^{k-1} c_j + c_{k+1} \sum_{j \geq k+1} c_j - c_k N - c_k^2.$$

   (b) Verify that the total density of clusters obeys $dN/dt = -c_1 N$.
   (c) Show that in the continuum limit the master equations reduce to

$$\frac{\partial c(k, t)}{\partial t} = -c_k(c_k + c_{k+1}) + N(c_{k-1} - c_k) + (c_{k+1} - c_{k-1}) \sum_{j \geq k} c_j$$

$$\simeq 2 \frac{\partial c}{\partial k} \left[\int_k^{\infty} dj\, c(j) \right] - N \frac{\partial c}{\partial k}.$$

   (d) Use the scaling ansatz $c_k \simeq N^2 \mathcal{C}(kN)$ to separate, in the continuum limit, the master equations into

$$\frac{dN}{dt} = -\mathcal{C}(0)N^3, \qquad \mathcal{C}(0)[2\mathcal{C} + x\mathcal{C}'] = 2\mathcal{C}^2 + \mathcal{C}' \left[1 - 2 \int_x^{\infty} dy \mathcal{C}(y) \right].$$

Solve these equations. *Hint:* Use the equation for $\mathcal{B}(x) = \int_0^x dy\, \mathcal{C}(y)$ to show that the scaled distribution $\mathcal{C}(x) = \mathcal{B}'(x)$ coincides with the zero-temperature Fermi distribution,
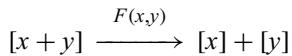
$$\mathcal{C}(x) = \begin{cases} \mathcal{C}(0), & x < x_*, \\ 0, & x \geq x_*. \end{cases}$$

Determine x_* using "mass" conservation.

Fragmentation

The continuous breakup of clusters, either by external driving or by mutual collisions, is known as *fragmentation*. At geological scales, fragmentation is responsible for sand grains on beaches and for boulder fields. Collision-driven fragmentation is responsible for "space junk," a potential long-term hazard for near-Earth satellites. At the molecular level, chemical bond breaking underlies polymer degradation and combustion. Moreover, simultaneous aggregation and fragmentation are responsible for interstellar dust clouds, planetary rings, and other heavenly structures; numerous biochemical reactions in cells provide more earthly examples. Fragmentation has the appearance of aggregation running backward in time and it is not surprising that these two processes are amenable to similar theoretical approaches. In this chapter we present the basic phenomenology of fragmentation and also examine the interplay between aggregation and fragmentation.

6.1 Binary fragmentation

Fragmentation may be represented as the reaction

$$[x + y] \xrightarrow{F(x,y)} [x] + [y]$$

in which a fragment of mass $x + y$ breaks into two smaller fragments of masses x and y with rate $F(x, y)$ that is necessarily symmetric in x and y.[1] In contrast to the complementary phenomenon of aggregation, it is natural to treat the cluster masses as continuous. In a close parallel with aggregation, our goal is to understand the emergent fragment size distribution. Since the number of fragments grows *ad infinitum* and the initial number of fragments is large in most physical settings, the simplest statistical framework that focuses on average densities and ignores fluctuations should give a progressively better approximation as the system evolves. The basic variable is thus $c(x, t)\, dx$, the density of fragments with mass in the range $(x, x + dx)$ at time t. This quantity evolves according to

$$\frac{\partial c(x, t)}{\partial t} = -c(x, t) \int_0^x F(y, x - y)\, dy + 2 \int_x^\infty c(y, t)\, F(x, y - x)\, dy. \tag{6.1a}$$

---

[1] The determination of the breakup rate $F(x, y)$ is a separate problem that depends on microscopic details of breakup events; here, we assume that these breakup rates are known. It is also possible to treat the breaking of an object into an arbitrary number of fragments by a master equation analysis.

The first term on the right accounts for the loss of fragments of mass x due to their breakup, while the second term accounts for the gain of such fragments due to the breakup of a cluster of mass $y > x$. The prefactor 2 accounts for the possibility that either one of the two daughter fragments from the breakup of a y-fragment has mass x. A useful check of the correctness of the master equation (6.1a) is to verify that the total mass, $\int x \, c(x, t) \, dx$, is constant.

When the breakup rate depends only on the initial mass and not on the masses of the daughter clusters, then $F(x, y) = \overline{F}(x + y)$ is a function of $x + y$ only. In this case, the master equation simplifies to

$$\frac{\partial c(x, t)}{\partial t} = -R(x) \, c(x, t) + 2 \int_x^\infty y^{-1} R(y) \, c(y, t) \, dy, \tag{6.1b}$$

where the total breakup rate of a fragment of size x is $R(x) = x\overline{F}(x)$.

The master equations of fragmentation (Eqs (6.1a) and (6.1b)) are based on a number of assumptions that are worthwhile to enumerate at the outset:

- Linearity: external driving causes fragmentation rather than interactions between fragments; thus the breakup rate of a fragment is independent of the state of other fragments.
- Binary fragmentation: two fragments are produced in each breaking event.
- Spatial homogeneity: fragment densities are independent of spatial position.
- Thermodynamic limit: the system is sufficiently large that discreteness effects may be ignored.
- Shape independence: fragment shape plays no role in the evolution.

Later in this chapter (Section 6.2), we will account for shape dependence in a simplistic way in a planar fragmentation model. We will also treat a simple model of nonlinear fragmentation in which collisions between fragments cause breakup events (Section 6.4).

Random scission

Perhaps the simplest example of fragmentation is *random scission* (Fig. 6.1), in which objects break at a rate that is proportional to their size. This interpretation is appropriate for
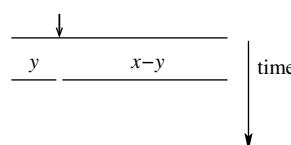
Fig. 6.1. Random scission. An initial segment of length x is cut at a random point into two segments of lengths y and $x - y$. The last line shows the result after many breaking events; equivalently, these can be viewed as a random deposition of cuts.

polymer degradation, where fragments are polymer chains and each polymer bond can break at the same fixed rate. In this case, we may define the total breaking rate to be $R(x) = x$. Then Eq. (6.1b) reduces to

$$\frac{\partial c(x,t)}{\partial t} = -xc(x,t) + 2\int_x^\infty c(y,t)\,dy. \tag{6.2}$$

Again, the first term on the right accounts for the loss of fragments of size x at a rate $R(x) = x$, while the second term accounts for the creation of two smaller fragments. Integrating (6.2) over all sizes, we see that the total number of fragments $N = \int c(x)\,dx$ grows linearly with time, $dN/dt = 1$ (with mass $\int x\,c(x,t)\,dx = 1$). Thus $N(t) = N(0) + t$, and the average fragment size $\langle x \rangle = 1/N$ asymptotically decays as
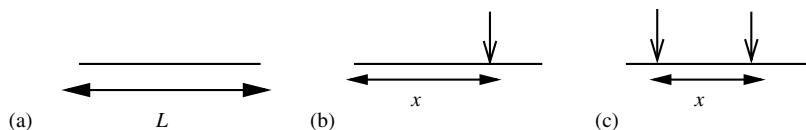
$$\langle x \rangle \simeq t^{-1}. \tag{6.3}$$

We start by giving a probabilistic explanation for the fragment size distribution. In this formulation, an interval is cut at random locations at a unit rate. The average number of cuts in an interval of length L thus equals Lt at time t. For an initial interval of length L, there will be three types of segments at some later time (Fig. 6.2):

- *Zero-cut segments.* An interval of length L contains no cuts with probability e^{-Lt}.
- *One-cut segments.* A segment whose length is in the range $(x, x + dx)$ is created when there are no cuts within a distance x from one end (with probability e^{-xt}) and the interval $(x, x + dx)$ contains a cut (with probability $t\,dx$). We need not care about the state of the remaining portion $L - x$. The full contribution of one-cut segments is $2t\,e^{-xt}$, because the segment can occur at either side.
- *Two-cut segments.* Such segments arise when there are two cuts in the intervals dy and dz at opposite ends of an otherwise intact interior segment of length x. The probability for this event is $t^2\,dy\,dz\,e^{-xt}$. We may again ignore the state of the portion exterior to the segment. Integrating over all possible segment positions gives $t^2(L - x)e^{-xt}$.

Combining these three contributions gives the probability $P_L(x, t)$ for a segment of length x at time t:

$$P_L(x,t) = e^{-xt}\left\{\delta(L-x) + \left[2t + t^2(L-x)\right]\Theta(L-x)\right\}. \tag{6.4}$$

(a) L (b)   (c)

The three types of segments of an interval: (a) zero-cut segment, (b) one-cut segment, and (c) two-cut segment.

We now integrate (6.4) over the initial length distribution, $c(x,t) = \int_0^\infty c_0(L) P_L(x,t)\, dL$, and obtain the size distribution

$$c(x,t) = e^{-xt} \left\{ c_0(x) + \int_x^\infty c_0(y) \left[2t + t^2(y-x) \right] dy \right\}. \tag{6.5}$$

In the long-time limit, the length distribution (6.5) approaches a scaling form. Out of the three terms in (6.5), only the last contributes to the scaling limit. Since the typical segment length decays to zero, we have

$$\int_x^\infty c_0(y)(y-x)\, dy \to \int_0^\infty c_0(y)\, y\, dy = 1,$$

and therefore the asymptotic mass distribution has a very simple scaling form:

$$c(x,t) \simeq t^2 e^{-xt}. \tag{6.6}$$

Homogeneous breakup rates

In many physical situations, the breakup rate is a homogeneous function of the fragment size. For example, in the random scission of a linear object, the breakup rate is proportional to its length. In situations where fragmentation is controlled by the surface area of an object, the breakup rate will grow as the $2/3$ power of the mass. As in aggregation, the primary parameter is the homogeneity index λ, which is defined by $R(ax) = a^\lambda R(x)$.[2] We now investigate fragmentation kinetics as a function of this homogeneity index. For the strict power-law breakup rate with $R(x) = x^\lambda$, the master equation (6.1b) is

$$\frac{\partial c(x,t)}{\partial t} = -x^\lambda\, c(x,t) + 2 \int_x^\infty y^{\lambda-1} c(y,t)\, dy. \tag{6.7}$$

The form of Eq. (6.7) suggests seeking a solution of the form

$$c(x,t) = \Psi(t)\, \exp\left[-\Phi(t)\, x^\lambda\right].$$

Substituting this ansatz into the master equation (6.7) gives

$$\frac{\dot\Psi}{\Psi} - x^\lambda \dot\Phi = -x^\lambda + \frac{2}{\lambda}\frac{1}{\Phi}.$$

Equating the coefficients of x^λ on both sides gives $\dot\Phi = 1$. Hence $\Phi(t)$ is a linear function of time, $\Phi = t + a$, where a is a constant. We then equate the remaining two terms and find

---

[2] Because fragments become progressively smaller, it is sufficient for homogeneity to hold only asymptotically; that is, when $x, y \to 0$.

$\Psi = (t+a)^{2/\lambda}$. Thus there is a family of exact solutions to (6.7) that is parameterized by a:

$$c(x,t) = (t+a)^{2/\lambda} \exp[-(t+a)x^{\lambda}]. \tag{6.8}$$

We now exploit the linearity of Eq. (6.7) and construct the general exact solution for $c(x,t)$ as a linear combination of elemental solutions:

$$c(x,t) = \int_0^{\infty} C(a)(t+a)^{2/\lambda} \exp[-(t+a)x^{\lambda}] \, da, \tag{6.9}$$

in which the amplitude $C(a)$ is determined from the initial condition. In the long-time limit, the solution (6.9) exhibits a universal behavior

$$c(x,t) \sim t^{2/\lambda} e^{-tx^{\lambda}}. \tag{6.10}$$

This fragment mass distribution can again be written in the standard scaling form $c(x,t) \sim \xi^{-2}f(x/\xi)$, with typical mass $\xi = t^{-1/\lambda}$ and scaling function $f(z) = e^{-z^{\lambda}}$.

Shattering

The preceding analysis tacitly assumed that $\lambda > 0$. When $\lambda < 0$, fragmentation of small particles becomes explosively fast, leading to the intriguing phenomenon of *shattering*, in which mass is lost to a "dust" phase of zero-mass particles. The shattering transition in fragmentation is the counterpart of the gelation transition in aggregation.

As in the case of gelation, the behavior of the moments provides the quickest way to detect the existence of a singularity. Integrating (6.7) over all x, we find that the moments

$$N(t) \equiv M_0(t) = \int_0^{\infty} c(x,t) \, dx, \qquad M_{\lambda}(t) = \int_0^{\infty} x^{\lambda} c(x,t) \, dx,$$

are related by

$$\frac{dN}{dt} = M_{\lambda}. \tag{6.11}$$

Equation (6.11) is exact and valid for all λ. To infer asymptotic behavior, it is simplest to follow a heuristic approach. In close analogy with our discussion of scaling in aggregation[3] let us exploit the feature that the fragment mass distribution has the scaling form, $c(x,t) = \xi^{-2}f(x/\xi)$. Using this form, the zeroth moment becomes

$$N = \int_0^{\infty} c(x,t) \, dx = \int_0^{\infty} \xi^{-2}f(x/\xi) \, dx \sim \xi^{-1}.$$

---

[3] See Section 5.4 and Eq. (5.56).

Similarly, the moment M_λ is

$$M_\lambda = \int_0^\infty x^\lambda\, c(x,t)\, dx \sim \xi^{\lambda-1}\,.$$

Combining these two relations gives $M_\lambda \sim N^{1-\lambda}$. Substituting this form for M_λ into Eq. (6.11) we obtain the closed equation for the density of fragments,

$$\frac{dN}{dt} \sim N^{1-\lambda}\,. \tag{6.12}$$

Hence the fragment density has the asymptotic time dependence

$$N \sim \begin{cases} t^{1/\lambda}, & \text{for } \lambda > 0, \\ e^{\omega t}, & \text{for } \lambda = 0, \\ (t_c - t)^{-1/|\lambda|}, & \text{for } \lambda < 0 \text{ and } t \to t_c. \end{cases} \tag{6.13}$$

When $\lambda < 0$, the fragmentation of the tiniest fragments occurs so quickly that the fragment density diverges at a finite time t_c. For $t > t_c$, a finite fraction of the mass is converted into a dust that consists of an infinite number of zero-mass particles that manages to contain a finite fraction of the total mass.

Fragmentation with input

Fragmentation with steady material input arises in many physical processes, such as crushing a steady stream of raw mineral ore. In such situations, the fragment size distribution attains a steady state. In fact, the master equation (6.7) generally admits a class of *steady-state* solutions of the form $c(x) = Cx^{-2-\lambda}$, in addition to the time-dependent solutions (6.8) (problem 6.4). Let us analyze the effect of input for homogeneous breakup rates. If the input rate of objects of mass x is $I(x)$, the mass distribution evolves according to

$$\frac{\partial c(x,t)}{\partial t} = -x^\lambda c(x,t) + 2\int_x^\infty y^{\lambda-1} c(y,t)\, dy + I(x). \tag{6.14}$$

We can solve this equation by employing the Mellin transform $c(s,t) \equiv \int_0^\infty x^{s-1}\, c(x,t)\, dx$. In terms of this transform, Eq. (6.14) becomes

$$\frac{\partial c(s,t)}{\partial t} = \frac{2-s}{s}\, c(s+\lambda,t) + I(s), \tag{6.15}$$

where $I(s) = \int_0^\infty x^{s-1} I(x)\, dx$ is the Mellin transform of the input rate. As in irreversible fragmentation, the Mellin transform again simplifies the original master equation into a soluble recursion formula.

The Mellin transform

The Mellin transform $f(s)$ of a function $f(x)$ and the corresponding inverse of the Mellin transform are given by

$$f(s) = \int_0^\infty x^{s-1} f(x)\, dx, \qquad f(x) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} x^{-s} f(s)\, ds. \qquad (6.16)$$

The notation implies that s should be treated as a complex variable and the integral in the inverse transform is taken over a vertical line in the complex plane. The Mellin transform is just the moment of order $s - 1$ of the function $f(x)$ and is also a Laplace transform in disguise: by defining $x = e^{-y}$, the Mellin transform turns into the Laplace transform. Thus we can adapt the known rules about inverting the Laplace transform to infer the inverse of the Mellin transform.

Here is a table of Mellin transforms of a few common functions:

$f(x)$	e^{-x}	$2e^{-x^2}$	$\ln(1+x)$	$(1+x)^{-1}$	$\Gamma(a)/(1+x)^a$	$(1+x^2)^{-1}$	$\sqrt{\pi}\,\mathrm{erfc}(x)$
$f(s)$	$\Gamma(s)$	$\Gamma(s/2)$	$\dfrac{\pi}{s\sin(\pi s)}$	$\dfrac{\pi}{\sin(\pi s)}$	$\Gamma(a-s)\,\Gamma(s)$	$\dfrac{\pi}{2\sin(\pi s/2)}$	$s^{-1}\Gamma[(1+s)/2]$

We shall often use two properties of the Mellin transform:

1. If $f(s)$ is the Mellin transform of $f(x)$, then $f(s - n)$ is the Mellin transform of $x^{-n} f(x)$.
2. The Mellin transform of $g(x) = \int_x^\infty y^{m-1} f(y)\, dy$ is $s^{-1} f(s + m)$. This relation becomes clear after an interchange of integration order:

$$\int_0^\infty x^{s-1} g(x)\, dx = \int_0^\infty x^{s-1}\, dx \int_x^\infty y^{m-1} f(y)\, dy$$

$$= \int_0^\infty y^{m-1} f(y)\, dy \int_0^y x^{s-1}\, dx$$

$$= \frac{1}{s} \int_0^\infty y^{s+m-1} f(y)\, dy = s^{-1} f(s + m)\,.$$

In the steady state, the Mellin transform in (6.15) satisfies (after shifting the index s by 1)

$$c(s) = \frac{s - \lambda}{s - 2 - \lambda}\, I(s - \lambda). \qquad (6.17)$$

We now rewrite the Mellin transform as

$$c(s) = \left(1 + \frac{2}{s - 2 - \lambda}\right) I(s - \lambda)$$

and use basic properties of the Mellin transform given in the highlight above to find that the steady-state fragment mass distribution is

$$c(x) = x^{-\lambda} I(x) + 2x^{-2-\lambda} \int_x^\infty I(y)\, y\, dy. \qquad (6.18)$$

In the limit of small fragment masses, the second term dominates so that the mass distribution has the universal algebraic tail,

$$c(x) \simeq \frac{2J}{x^{2+\lambda}} \qquad \text{as} \quad x \to 0, \tag{6.19}$$

that was mentioned at the outset of this section; here $J = \int_0^\infty x\,I(x)\,dx$ is the injected density of material of mass x per unit time. Thus apart from an overall amplitude J, the stationary mass distribution is *independent* of the details of the input distribution.

This asymptotic behavior can also be obtained by the following heuristic argument. Let's consider the linear fragmentation rate $R(x) = x$ for concreteness. Because of the steady input, the total mass in the system, $M(t) = c(s = 2, t)$, grows linearly with time, $M(t) = Jt$. Similarly, from Eq. (6.14), the total number of fragments $N(t) = c(s = 1, t)$ satisfies $\dot{N}(t) = Jt + \mu$, where $\mu = \int I(x)dx$ is the number of fragments added per unit time. Consequently, $N(t) = \frac{1}{2}Jt^2 + \mu t$. The first two moments imply that the typical fragment mass is $M/N \sim t^{-1}$. Thus the mass distribution should approach the scaling form

$$c(x, t) \simeq t^3 F(xt) \qquad \text{as} \quad t \to \infty. \tag{6.20}$$

The prefactor t^3 ensures that the total mass in the system grows linearly with time, while the argument in the scaling function gives the $1/t$ decay of the typical fragment mass. With this scaling form, a steady state is possible only when $F(z) \sim z^{-3}$ for large z. This fact implies that $c(x) \sim x^{-3}$.

Example 6.1. *Charlesby method for fragmentation with input.* For the full time-dependent solution for aggregation with input, we start by expanding the Mellin transform as an infinite power series in time[4]

$$c(s, t) = \sum_{k \geq 1} \frac{t^k}{k!} M_k(s), \tag{6.21}$$

and then determine the expansion functions $c_k(s)$ iteratively. For an initially empty system, this expansion does not contain a constant term, while for a steady input, the Mellin transform of the input contains only a time-independent term $I(s)$.

We now substitute the expansion (6.21) into Eq. (6.15) and equate terms with the same powers of time. This yields $M_1(s) = I(s)$, and $M_{k+1}(s) = -(1 - 2/s)\,M_k(s+1)$ for $k \geq 2$. Solving this set of equations recursively gives

$$M_{k+1}(s) = (-1)^k \frac{(s-1)(s-2)}{(s+k-1)(s+k-2)} I(s+k).$$

---

[4] This method is rarely used as it looks like magic. Indeed, the method relies on a Taylor expansion in time that is usually useful at short times, but we are actually interested in the long-time behavior. The trick works when only a finite number of terms in the Taylor series remains non-zero. The method originates with Euler; in the context of fragmentation, this technique was used by Charlesby.

To invert this Mellin transform, we rewrite $M_{k+1}(s)$ as the partial fraction expansion

$$M_{k+1}(s) = (-1)^k \left[1 - \frac{k(k+1)}{s+k-1} + \frac{k(k-1)}{s+k-2} \right] I(s+k). \tag{6.22}$$

Using Eq. (6.21) we express the mass distribution as a power series

$$c(x,t) = \sum_{k \geq 0} \frac{t^{k+1}(-x)^k}{(k+1)!} c_k(x), \tag{6.23}$$

where the inverse transform of $M_{k+1}(s)$ has been conveniently written as $(-x)^k c_k(x)$. The three terms in (6.22) can be inverted using the rules outlined in the Mellin transform highlight on page 177 to yield:

$$c_k(x) = I_1(x) - \frac{k(k+1)}{x} I_2(x) + \frac{k(k-1)}{x^2} I_3(x), \tag{6.24}$$

with

$$I_1(x) = I(x), \qquad I_2(x) = \int_x^\infty I(y)\,dy, \qquad I_3(x) = \int_x^\infty y\,I(y)\,dy. \tag{6.25}$$

Equation (6.24) allows us to transform the infinite series (6.23) into three infinite subseries that are readily summable. This gives the mass distribution as the sum of just three terms,

$$c(x,t) = \sum_{k=1}^{3} t^k I_k(x) F_k(xt), \tag{6.26}$$

with the scaling functions[5]

$$\begin{aligned}
F_1(z) &= z^{-1}(1 - e^{-z}), \\
F_2(z) &= e^{-z}, \\
F_3(z) &= z^{-3}\left[2 - (2 + 2z + z^2)e^{-z} \right].
\end{aligned} \tag{6.27}$$

Equations (6.26) and (6.27) give the full time-dependent solution for an *arbitrary* steady input $I(x)$. In the limit $x \to 0$ and $t \to \infty$, with the scaling variable $z = xt$ kept finite, the third term in the sum of Eq. (6.26) dominates, and the anticipated scaling behavior of (6.20) is recovered, with $F(z) = F_3(z)$.

6.2 Planar fragmentation

In our presentation of fragmentation kinetics thus far, a single factor – the fragment size – has controlled the kinetics. However, geometric aspects are clearly relevant, as evidenced by our

---

[5] The function $F_3(z)$ has been obtained from the power series $F_3(z) = \sum_{k \geq 0} (-z)^k / [k!(k+3)]$.
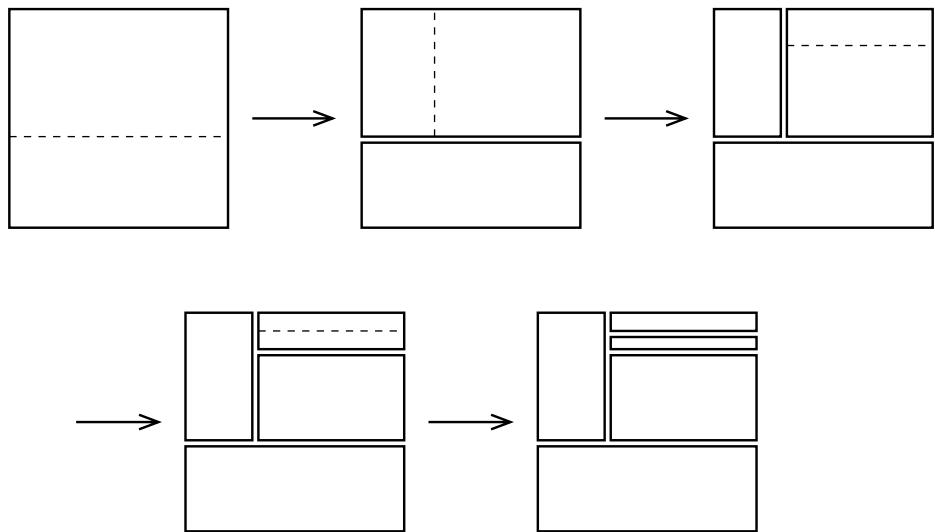
Fig. 6.3.   **First few steps in rectangular fragmentation.**
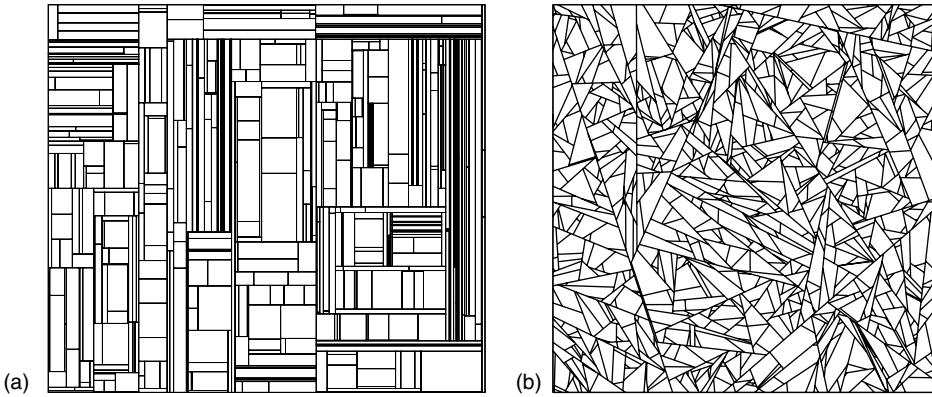
everyday experience with the breaking of brittle objects. Here we discuss a fragmentation model that incorporates the effects of shape in a minimalist way – rectangular fragmentation – in which a single rectangle is fragmented into many smaller rectangles by a series of straight cuts. A single update event in rectangular fragmentation consists of the following (Fig. 6.3):

1. pick a random rectangle with a probability that is proportional to its area;
2. from a random point inside the selected rectangle, a crack through this point – either horizontal or vertical with equal probability – fragments the rectangle into two smaller rectangles in an area-preserving manner.[6]

Models of this spirit have been proposed to describe the formation of martensites in two dimensions, where the refinement of microdomains in martensites visually resembles the outcome after many rectangular fragmentation events. (Fig. 6.4(a) shows the fragmentation of a unit square after 500 such breaking events.) A striking feature of rectangular fragmentation is that it satisfies an infinite number of "hidden" conservation laws. As a consequence, the mass distribution obeys *multifractal* scaling, rather than the conventional scaling that arises in kinetic fragmentation and aggregation models.

Without loss of generality, we assume that the system starts with a single 1×1 square and we set the total fragmentation rate to 1. Since one rectangle is destroyed and two daughter rectangles are created in each breaking event, the average number of fragments grows linearly in time: $N = 1 + t$. The population of rectangles is characterized by the probability distribution $c(x_1, x_2, t)$, in which $c(x_1, x_2, t)\, dx_1\, dx_2$ is the density of rectangles with length

---

[6] It is straightforward to generalize to arbitrary spatial dimension d, where a $(d-1)$-dimensional hyperplane grows in one of $d-1$ possible orthogonal directions from a given seed.

Fig. 6.4. Long-time outcomes of planar fragmentation of the unit square. (a) Rectangular fragmentation after 500 breaking events, (b) isotropic fragmentation after 1000 breaking events.

and width in the range $(x_1, x_1 + dx_1)$ and $(x_2, x_2 + dx_2)$ at time t. This distribution evolves according to

$$\frac{\partial c(x_1, x_2, t)}{\partial t} = -x_1 x_2 \, c(x_1, x_2, t) + x_2 \int_{x_1}^1 c(y_1, x_2, t) \, dy_1 + x_1 \int_{x_2}^1 c(x_1, y_2, t) \, dy_2. \quad (6.28)$$

The first term on the right accounts for the loss of a rectangle of area $A = x_1 x_2$ due to its fragmentation. Since the breaking point is chosen randomly inside the total area of the initial rectangle, the overall breaking rate of any rectangle is proportional to its area. The second term accounts for the gain of a fragment of length x_1 and height x_2 by vertically cracking a rectangle of length $y_1 > x_1$ and height x_2. The prefactor x_2 accounts for the breaking point being situated anywhere along the vertical crack. The last term similarly accounts for the contribution due to a horizontal crack.

To simplify (6.28), we employ the two-variable Mellin transform

$$c(s_1, s_2, t) = \int_0^1 \int_0^1 x_1^{s_1-1} x_2^{s_2-1} \, c(x_1, x_2, t) \, dx_1 \, dx_2 \, .$$

Multiplying both sides of the master equation by $x_1^{s_1-1} x_2^{s_2-1}$ and integrating over all x_1 and x_2 we find that the Mellin transform obeys the linear function equation

$$\frac{\partial c(s_1, s_2, t)}{\partial t} = \left(\frac{1}{s_1} + \frac{1}{s_2} - 1 \right) c(s_1 + 1, s_2 + 1, t) \, . \quad (6.29)$$

Rather than attempting to solve this equation directly, let's study the moments of the fragment mass distribution. From the definition of the Mellin transform, these moments are given by

$$\langle x_1^{n_1} x_2^{n_2} \rangle \equiv \frac{c(n_1 + 1, n_2 + 1)}{c(1, 1)} \, . \quad (6.30)$$

We know that conservation of the total area of all fragments gives $c(2,2,t) = 1$. We also know the average number of fragments is given by $c(1,1,t) = 1 + t$. Consequently, the average area of a fragment is $\langle A \rangle = \langle x_1 x_2 \rangle = 1/(1+t)$.

A striking feature of the equation of motion (6.29) is that it contains infinitely many conservation laws that are defined by the curve in the s_1–s_2 plane[7]

$$(s_1^*)^{-1} + (s_2^*)^{-1} = 1 \qquad\qquad (6.31)$$

along which the Mellin transform $c(s_1^*, s_2^*)$ is constant in time. It is important to emphasize that these hidden conservation laws hold only in an average sense. That is, while $c(s_1^*, s_2^*)$ does not remain constant in each individual breaking event, $c(s_1^*, s_2^*)$, averaged over all realizations of rectangular fragmentation, is conserved. The only strict conservation law is that the total area never changes in any fragmentation event. The curve (6.31) can be represented in the parametric form $s_1^* = \kappa + 1, s_2^* = \kappa^{-1} + 1$. The corresponding conservation laws $c(s_1^*, s_2^*, t) = c(s_1^*, s_2^*, 0)$ show that the moments of the form

$$\int_0^1 \int_0^1 x_1^\kappa \, x_2^{1/\kappa} \, c(x_1, x_2, t) \, dx_1 \, dx_2$$

are conserved for all $\kappa > -1$. The case $\kappa = 1$ corresponds to area conservation; other κ values, even the simplest ones such as $\kappa = 2$, lead to conserved moments that lack a clean physical meaning.

Since the average mass decays algebraically with time in linear fragmentation with homogeneous breakup kernels, it is natural to assume that the Mellin transform also decays algebraically with time for rectangular fragmentation; that is, $c(s_1, s_2) \sim t^{-\alpha(s_1, s_2)}$, with unknown exponent $\alpha(s_1, s_2)$. Substituting this form into (6.29) immediately leads to the recursion for the exponent $\alpha(s_1 + 1, s_2 + 1) = \alpha(s_1, s_2) + 1$. Combining this recursion with the conservation law $\alpha(s_1^*, s_2^*) = 0$, we determine the exponent $\alpha(s_1, s_2)$ to be

$$\alpha(s_1^* + k, s_2^* + k) = k \qquad \text{for all} \qquad (s_1^*)^{-1} + (s_2^*)^{-1} = 1. \qquad (6.32)$$
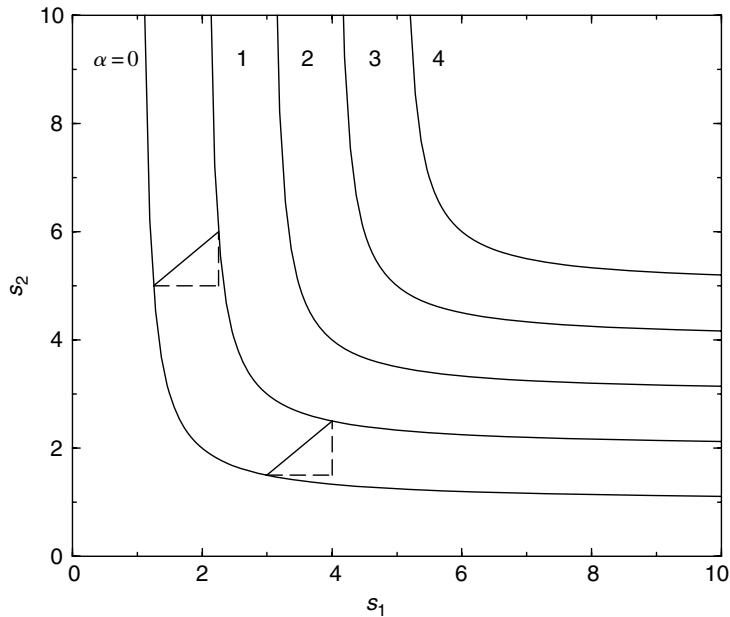
Pictorially, the value of α at an arbitrary point (s_1, s_2) is just the horizontal (or vertical) distance from this point to the curve $(s_1)^{-1} + (s_2)^{-1} = 1$ (Fig. 6.5). This distance condition gives $(s_1 - \alpha) = s_1^*$ and $(s_2 - \alpha) = s_2^* = [1 - (s_1^*)^{-1}]^{-1}$. Eliminating s_1^* from these two equations, the exponent is the smaller root of the quadratic equation $(s_1 - \alpha)(s_2 - \alpha) - (s_1 - \alpha) - (s_2 - \alpha) = 0$, which gives

$$\alpha(s_1, s_2) = \frac{s_1 + s_2}{2} - 1 - \sqrt{\frac{(s_1 - s_2)^2}{4} + 1} \ .$$

Using these exponent relations, the asymptotic behaviors of the moments (6.30) are given by

$$\langle x_1^{s_1} x_2^{s_2} \rangle \sim t^{-\alpha(s_1+1, s_2+1) + \alpha(1,1)} \sim t^{-\alpha(s_1, s_2) - 2} \ . \qquad (6.33)$$

---

[7] Rewriting (6.31) as $(s_1^* - 1)(s_2^* - 1) = 1$ we see that geometrically the curve (6.31) is a hyperbola.

Fig. 6.5. Loci of $\alpha(s_1, s_2) = 0, 1, 2, 3,$ and 4 from Eq. (6.32). Each successive curve is shifted horizontally and vertically by 1 from its predecessor (dashed lines).

As a specific example, the moments of the fragment length are given by

$$\langle \ell^n \rangle \equiv \langle x_1^n \rangle \sim t^{-(n+2-\sqrt{n^2+4})/2}, \tag{6.34}$$

so that $\langle \ell \rangle \sim t^{-(3-\sqrt{5})/2} \sim t^{-0.382}$ and $\langle \ell^2 \rangle^{1/2} \sim t^{-(2-\sqrt{2})/2} \sim t^{-0.293}$. These moments decay more slowly than what one might naively anticipate from the behavior of the average area. Since $\langle A \rangle = (1 + t)^{-1}$, the natural length scale should ostensibly be $t^{-1/2}$, which decays much faster than $\langle \ell \rangle$. Another intriguing property of rectangular fragmentation is that a typical fragment becomes more elongated as a function of time Figs 6.3 and 6.4(a). We quantify this asymmetry by the aspect ratio moments $\langle (x_1/x_2)^n \rangle$ which, from Eq. (6.33), diverge as $\langle (x_1/x_2)^n \rangle \sim t^{\sqrt{n^2+1}-1}$ in the long-time limit.

The nonlinear spectrum of scaling exponents $\alpha(s_1, s_2)$ represents an example of *multiscaling*, in which high-order moments are not directly related to low-order moments. For rectangular fragmentation, the consequences of this multiscaling are nicely illustrated by the anomalous properties of the length moments and of the aspect ratio. It bears emphasizing that the restriction to oriented cracks is an important ingredient in this multiscaling. If cracks are isotropic, then conventional scaling is restored, as is visually apparent in the example of isotropic planar fragmentation in Fig. 6.4(b). In this isotropic fragmentation, a fragment is chosen with rate proportional to its area and is then cut by a line of random position and orientation. For this model, the characteristic linear dimension ℓ and area A of fragments now obey the conventional scaling $\langle \ell \rangle^2 \sim \langle A \rangle$.

Example 6.2. *Area distribution*. Denote by $F(A, t)$ the density of rectangles of area A. This area distribution can be expressed in terms of $c(x_1, x_2, t)$ by

$$F(A, t) = \int \int c(x_1, x_2, t)\, \delta(x_1 x_2 - A)\, dx_1\, dx_2 . \tag{6.35}$$

Equation (6.35) is not very useful for concrete calculations, as we do not have an explicit expression for $c(x_1, x_2, t)$. However, we can determine the area distribution by using its Mellin transform $F(s, t)$. From Eq. (6.35), we have $F(s, t) = c(s, s, t)$. Hence, the equation of motion for $F(s, t)$ is just the diagonal form of Eq. (6.29), that is

$$\frac{\partial F(s, t)}{\partial t} = \left(\frac{2}{s} - 1\right) F(s + 1, t) . \tag{6.36}$$

As in the case of the length moments, we assume that $c(s, s) = \mu(s) t^{-\alpha(s)}$ and substitute into Eq. (6.36) to give $\alpha(s) = s - 2$. This equation also gives the relation for the amplitude $\mu(s + 1) = s\,\mu(s)$. Using the boundary values $\mu(1) = 1$ (normalization) and $\mu(2) = 1$ (total area equals 1), we obtain $\mu(s + 1) = s\mu(s)$, and so we finally obtain

$$F(s, t) \simeq \Gamma(s)\, t^{2-s} \qquad \text{when} \quad t \gg 1. \tag{6.37}$$
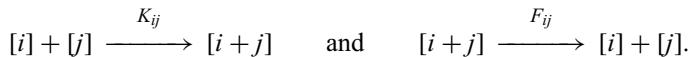
Inverting this Mellin transform we get

$$F(A, t) \simeq t^2\, e^{-At} . \tag{6.38}$$

The scaled area distribution has the same form as the length distribution in the random scission model.

6.3 Reversible polymerization

In polymer science, aggregation and fragmentation often proceed simultaneously to give reversible polymerization. Consider a population of polymer chains that are built from identical monomers that each contain two reactive endgroups. Symbolically, each chain can be represented as –A–A–$\cdots$–A–A–. Bonding leads to two polymers of lengths i and j aggregating into a longer polymer with rate K_{ij}, while the breakup into two shorter polymers of lengths i and j proceeds with rate F_{ij}:

$$[i] + [j] \xrightarrow{K_{ij}} [i + j] \qquad \text{and} \qquad [i + j] \xrightarrow{F_{ij}} [i] + [j].$$

The aggregation and fragmentation rates are (semi-)infinite symmetric matrices, $K_{ij} = K_{ji}$ and $F_{ij} = F_{ji}$.

The density of polymers that contain k monomers, $c_k(t)$, evolves by the master equations

$$\frac{dc_k}{dt} = \frac{1}{2} \sum_{i+j=k} K_{ij}\, c_i\, c_j - c_k \sum_{j \geq 1} K_{kj}\, c_j + \sum_{j \geq 1} F_{kj} c_{j+k} - \frac{1}{2}\, c_k \sum_{i+j=k} F_{ij} . \tag{6.39}$$

The mass distribution typically becomes stationary due to the competition between aggregation and fragmentation. In the simplest situations, the stationary distribution may be found by equating the rate of the aggregation process $[i] + [j] \rightarrow [i+j]$ to that of the reverse fragmentation process $[i+j] \rightarrow [i] + [j]$. This *detailed balance condition* (see the highlight below),

$$K_{ij} \, c_i \, c_j = F_{ij} \, c_{i+j} \,, \tag{6.40}$$

determines the equilibrium distribution.

Because the relaxation to equilibrium is usually exponentially quick and can often be ignored, we only need to study the equilibrium system of algebraic equations (6.40) rather than the system of differential equations (6.39). This simplification evokes a lesson from kinetic theory (Chapter 3) – the governing Boltzmann equation is intimidating and essentially no analytical solutions are known, but the corresponding analog of (6.40) has a solution, i.e. the Maxwell–Boltzmann distribution. There is, however, a crucial difference between reversible polymerization and kinetic theory: Eqs (6.40) may have no solutions.

To appreciate this point, let's write the first few of Eqs (6.40):

$$\begin{aligned}
K_{11} \, c_1^2 &= F_{11} \, c_2, \\
K_{12} \, c_1 c_2 &= F_{12} \, c_3, \\
K_{13} \, c_1 c_3 &= F_{13} \, c_4, \\
K_{22} \, c_2^2 &= F_{22} \, c_4.
\end{aligned} \tag{6.41}$$

We see that c_2 and c_3 admit unique solutions in terms of c_1:

$$c_2 = \frac{K_{11}}{F_{11}} \, c_1^2, \qquad c_3 = \frac{K_{11}}{F_{11}} \frac{K_{12}}{F_{12}} \, c_1^3 \,.$$

Substituting these results into the third and fourth of Eqs (6.41) we obtain *two independent* expressions for c_4:

$$c_4 = \frac{K_{11}}{F_{11}} \frac{K_{12}}{F_{12}} \frac{K_{13}}{F_{13}} \, c_1^4, \qquad c_4 = \left(\frac{K_{11}}{F_{11}} \right)^2 \frac{K_{22}}{F_{22}} \, c_1^4 \,. \tag{6.42}$$

If the constraint

$$\frac{K_{12}}{F_{12}} \frac{K_{13}}{F_{13}} = \frac{K_{11}}{F_{11}} \frac{K_{22}}{F_{22}} \tag{6.43}$$

is satisfied, Eqs (6.42) are consistent; otherwise there is no solution. Therefore Eqs (6.40) generally form an *overdetermined* system – there are infinitely many constraints like (6.43), and the solution set is usually empty. When many of the reaction rates vanish, the system of equations (6.40) may be unsolvable, or this system may become *underdetermined*, that is, it may have many solutions. Even if Eqs (6.40) admit a unique solution, we should establish that this solution is actually an *attracting* fixed point and that the initial condition belongs to its basin of attraction.

Detailed balance

The detailed balance condition differentiates two types of stationary states. In general, the flux $J_{i\to j}$ from state i to state j equals the product of p_i, the probability the system is in state i, times $w_{i\to j}$, the transition rate from state i to state j. That is $J_{i\to j} = p_i w_{i\to j}$. For a steady state, the total flux into state i must equal the total flux out of that state:

$$\sum_{j\neq i} p_i w_{i\to j} = \sum_{j\neq i} p_j w_{j\to i}, \tag{6.44}$$

for all i. On the other hand, detailed balance requires that the fluxes between *any* two states balance; that is, $J_{i\to j} = J_{j\to i}$, or
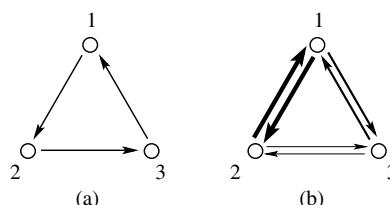
$$p_i w_{i\to j} = p_j w_{j\to i} \quad \text{for all } i,j. \tag{6.45}$$

Clearly, detailed balance (6.45) guarantees that the steady-state condition (6.44) is satisfied, but steady state does not necessarily imply detailed balance. A simple example of a non-equilibrium steady state is given in Fig. 6.6(a), where only cyclic transitions are allowed, $w_{1\to 2}, w_{2\to 3}, w_{3\to 1} \neq 0$, while all other transitions cannot occur. The steady-state flux $J = p_1 w_{1\to 2} = p_2 w_{2\to 3} = p_3 w_{3\to 1}$ satisfies (6.44) but violates detailed balance (6.45) (problem 6.7). For a system with N states, there are N steady-state conditions, but a much larger number, $\binom{N}{2} = \frac{1}{2}N(N-1)$, of detailed balance conditions.

In general, there are two types of stationary distributions in combined aggregation and fragmentation: (i) equilibrium steady states that satisfy the detailed balance condition (6.45) and (ii) non-equilibrium steady states that satisfy the master equation (6.44) but violate detailed balance. The latter case involves non-vanishing net flux between states. We now look more carefully at two representative models that illustrate this distinction.

String model

A classic description for the evolution of linear polymerization at finite temperature is the *string model*, in which the rates of aggregation and fragmentation are assumed to be constant. This model is based on the reasonable assumption that the reactivities of all chemical bonds

Fig. 6.6. (a) Non-equilibrium steady state with cyclic flux; (b) an equilibrium steady state with detailed balance.

are identical. For this model, Eqs (6.40) become

$$c_i \, c_j = \lambda \, c_{i+j}, \tag{6.46}$$

where $\lambda = F_{ij}/K_{ij}$ is the ratio of fragmentation and aggregation rates. Equations (6.46) admit a one-parameter family of solutions

$$c_k = \lambda \, \beta^k , \tag{6.47}$$

with the value of β fixed by the condition that the total mass, $\sum_{k \geq 1} k c_k$, equals 1. This condition gives the quadratic equation $\lambda \beta /(1 - \beta)^2 = 1$, with the proper root

$$\beta = 1 + \frac{\lambda}{2} - \sqrt{\lambda + \frac{\lambda^2}{4}} , \tag{6.48}$$

that is dictated by the requirement that the distribution is a decaying function of k.

Example 6.3. *Equilibrium polymerization*. Let's verify that the asymptotic solution of the full time-dependent master equations is indeed the equilibrium solution, Eqs (6.47)–(6.48). We choose $K_{ij} = 2$ and $F_{ij} = 2\lambda$, so that the time-dependent master equations (6.39) become

$$\dot{c}_k = \sum_{i+j=k} c_i c_j - 2 c_k M_0 + 2\lambda \sum_{j>k} c_j - \lambda (k-1) c_k , \tag{6.49}$$

where $M_0 = \sum_{k \geq 1} c_k$, the zeroth moment of the mass distribution, is the cluster density. To solve Eqs (6.49), we use the exponential ansatz

$$c_k = (1 - a)^2 \, a^{k-1}$$
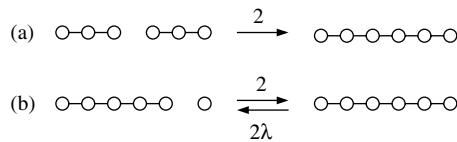
that manifestly satisfies mass conservation, $\sum_{k \geq 1} k c_k = 1$. For the monomer-only initial state, $a(t=0) = 0$. The exponential ansatz recasts the infinite set of differential equations (6.49) into a single differential equation $\dot{a} = (1 - a)^2 - \lambda a$. The solution is

$$a(t) = \frac{1 - e^{-(\alpha - \beta)t}}{\alpha - \beta \, e^{-(\alpha - \beta)t}} , \qquad \text{with} \quad \alpha = 1 + \frac{\lambda}{2} + \sqrt{\lambda + \frac{\lambda^2}{4}} . \tag{6.50}$$

Since $a(t = \infty) = \alpha^{-1} = \beta$, the resulting stationary mass distribution is $c_k(t = \infty) = (1 - \beta)^2 \, \beta^{k-1} = \lambda \, \beta^k$, which agrees with (6.47)–(6.48).

Chipping model

As an example of a non-equilibrium steady state, consider the situation where polymer growth competes with degradation – or chipping – at the end of the chain. Since end monomers are attached to a polymer by a single bond while the interior monomers are

The constituent proccess in the chipping model: (a) irreversible aggregation and (b) reversible monomer chipping. The rates of each process are indicated.

attached by two bonds, end monomers are more susceptible to detachment. The chipping model represents the extreme case where only end monomers can detach (Fig. 6.7):

$$F_{ij} = \begin{cases} 2\lambda, & i = 1 \text{ or } j = 1, \\ 0, & i \neq 1 \text{ and } j \neq 1. \end{cases} \tag{6.51}$$

We treat the simplest constant-kernel aggregation process ($K_{ij} = 2$) that naturally applies to polymerization. The chipping rates, $[k + 1] \rightarrow [k] + [1]$, are also assumed to be mass independent and we denote these rates by 2λ. Detailed balance is clearly violated for the chipping model because not all microscopic steps are reversible. As we now discuss, the chipping model leads to an unusual polymer length distribution in which a critical phase coexists with a gel phase.

The densities of polymers of length k, $c_k(t)$, evolve according to

$$\frac{dc_k}{dt} = -2c_k N + \sum_{i+j=k} c_i c_j + 2\lambda(c_{k+1} - c_k), \quad k \geq 2, \tag{6.52a}$$

$$\frac{dc_1}{dt} = -2c_1 N + 2\lambda\left(c_2 + \sum_{j \geq 2} c_j\right), \tag{6.52b}$$

where $N(t) = \sum_{j \geq 1} c_j(t)$ is the total polymer density. As should now be familiar, we solve these master equations by introducing the generating function $\mathcal{C}(z, t) = \sum_{j \geq 1}(z^j - 1) c_j(t)$, to recast Eqs (6.52a)–(6.52b) into the Riccati equation

$$\frac{\partial \mathcal{C}}{\partial t} = \mathcal{C}^2 + 2\lambda \frac{1 - z}{z} \mathcal{C} + 2\lambda \frac{(1 - z)^2}{z} N. \tag{6.53}$$

In the stationary state ($t \rightarrow \infty$), Eq. (6.53) reduces to a quadratic equation whose solution is

$$\mathcal{C}(z) = \lambda\left(1 - z^{-1}\right)\left(1 - \sqrt{1 - 2Nz/\lambda}\right). \tag{6.54}$$

We obtain the mass density by differentiating (6.54) with respect to z and setting $z = 1$. Since the initial mass density equals one, we obtain

$$1 = \sum_{j \geq 1} j c_j = \frac{d\mathcal{C}}{dz}\bigg|_{z=1} = \lambda\left(1 - \sqrt{1 - 2N/\lambda}\right), \tag{6.55}$$

whose solution is $N = 1 - (2\lambda)^{-1}$.

This solution cannot be correct for $\lambda < 1/2$ because the polymer density would be negative. In fact, Eq. (6.55) is inconsistent already for $\lambda < 1$ because the factor inside the square root becomes negative. Physically this inconsistency is caused by the appearance of an infinite polymer ("gel"). Mathematically, the sum $\sum_{j\geq1} j\, c_j$ that accounts for all finite polymers ("sol") becomes smaller than one when $\lambda < 1$; the gel contains the remaining fraction $g = 1 - \sum_{j\geq1} j\, c_j$ of the initial mass. From Eq. (6.54) we deduce that $N = \lambda/2$ in the gel phase (problem 6.10). In summary, the total polymer density is given by

$$N = \begin{cases} 1 - (2\lambda)^{-1}, & \lambda \geq 1 \quad \text{(sol phase)}, \\ \lambda/2, & \lambda < 1 \quad \text{(gel phase)}. \end{cases} \tag{6.56}$$

This change in behavior indicates that the system undergoes a non-equilibrium phase transition at $\lambda = 1$. Since N is continuous through the critical point $\lambda_c = 1$, the phase transition is continuous.

In the gel phase, the generating function (6.54) becomes

$$\mathcal{C}(z) = \lambda \left(1 - z^{-1} \right) \left(1 - \sqrt{1 - z} \right). \tag{6.57}$$

Differentiating (6.57) with respect to z and setting $z = 1$, we find that the sol mass $\sum_{j\geq1} j\, c_j = \lambda$. Therefore the mass of the gel is

$$g = \begin{cases} 0, & \lambda \geq 1, \\ 1 - \lambda, & \lambda < 1. \end{cases} \tag{6.58}$$

This behavior is another manifestation of a continuous phase transition in which the mass of the gel plays a role analogous to the order parameter in ordinary (equilibrium) continuous phase transitions.

While the properties outlined above suggest that the system undergoes a conventional continuous phase transition, it is more natural to think of the gel as a kind of critical phase for all $\lambda \leq 1$. To illustrate this point, we study the distribution of polymer lengths by expanding (6.54) in powers of z. In the sol phase ($\lambda \geq 1$), we obtain

$$c_k = e^{-\Lambda k} \, \frac{\Gamma\left(k - \frac{1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma(k+1)} \left[\frac{(\lambda - 1)^2}{2\lambda} + \frac{3}{2}\frac{1 - (2\lambda)^{-1}}{k+1} \right], \tag{6.59}$$

with $\Lambda = 2\ln\lambda - \ln(2\lambda - 1)$. Asymptotically the density of polymers of length $k \gg 1$ decays exponentially with k, but with an algebraic $k^{-3/2}$ prefactor:

$$c_k \simeq C\, k^{-3/2}\, e^{-\Lambda k}, \qquad C = \frac{(\lambda - 1)^2}{\lambda\sqrt{4\pi}}. \tag{6.60}$$

This result is not surprising – away from the critical point the behavior should be non-critical. However, for the gel phase we expand (6.57) and obtain

$$c_k = \frac{3\lambda}{4\sqrt{\pi}} \frac{\Gamma\left(k - \frac{1}{2}\right)}{\Gamma(k+2)} \simeq \frac{3\lambda}{4\sqrt{\pi}} k^{-5/2} \qquad \text{when} \quad k \gg 1. \tag{6.61}$$

Remarkably, the polymer mass distribution is algebraic, not only at the critical point $\lambda_c = 1$, but in the entire regime $\lambda \le \lambda_c = 1$! This criticality runs counter to the common lore of continuous phase transitions, in which critical behavior should arise only at the critical point. In the regime $\lambda < 1$, the excess mass (namely the fraction $1 - \lambda$ of the total mass of the system) condenses into a giant polymer (the gel), while the rest of the system remains in the critical state.

6.4 Collisional fragmentation

In astrophysics and materials science, collisions between distinct clusters can lead to fragmentation (Fig. 6.8). Such a process may be represented as

$$[x] + [y + z] \longrightarrow [x] + [y] + [z],$$
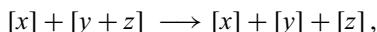
Fig. 6.8.   Collision-induced fragmentation. This image is an artist's rendering of the environment around Pleiades star HD 23514. (Image produced by Lynette R. Cook and reproduced with permission from Gemini Observatory/AURA.)

in which the collision of two fragments leads to breaking of one of the fragments so that the net outcome is three fragments. Collision-induced fragmentation is likely an intense process and a breakup event often results in many fragments. Nevertheless, the violent nature of collision-induced fragmentation manifests itself even in the mildest framework of binary breakup events.

Let's start with the most general features of collisional fragmentation. Since breakup events are induced by collisions, they occur with a rate proportional to the square[8] of the fragment density, N^2. Thus in suitable units, N evolves according to

$$\frac{dN}{dt} = N^2 , \tag{6.62}$$

with solution

$$N(t) = \frac{1}{1-t} , \tag{6.63}$$

when the initial density equals 1. Thus the number of fragments diverges at a finite shattering time $t_c = 1$, where the average fragment mass vanishes. While shattering can also occur in linear fragmentation, this feature arises only for pathological breakup kernels in which small particles break more easily than the larger ones; in reality, the opposite usually holds. However, in collision-induced fragmentation, shattering arises generically for any type of breakup rules.

To determine the mass distribution in collisional fragmentation, we will construct idealized models in which the basic ingredients include the following:

1. *What breaks*. Natural rules include (i) a randomly chosen particle splits upon collision, or (ii) the larger particle splits.
2. *How fragments break.* We assume deterministic breakup in which the fragment that breaks always splits into two equal-mass pieces. If initially all particles have unit mass, a fragment has mass $m = 2^{-n}$ after n breakup events.[9]

For fragments that split into two equal pieces, we write $c_n(t)$ for the density of fragments that have undergone n splitting events up to time t and thus have mass 2^{-n}.

Example 6.4. *Random particle splits*. Suppose that a randomly selected particle splits upon collision (equivalent to having both particles split at one-half the rate). The density $c_n(t)$ evolves according to

$$\frac{dc_n}{dt} = N\left[2c_{n-1} - c_n\right], \tag{6.64}$$

with the total density $N(t) = \sum_{j \geq 0} c_j(t)$. The factor 2 in the gain term arises because the splitting of a particle of mass $2^{-(n-1)}$ creates two particles of mass 2^{-n}. As a useful

---

[8] This statement ignores the decreasing fragment size; we treat this more physical situation below.
[9] The more physical rule that the splitting is stochastic leads to substantially similar behavior to that in deterministic splitting; see problem 6.14.

consistency check, summing Eqs (6.64) for all n recovers the rate equation (6.62) for the density of clusters of any size. Similarly, one may verify that the mass density, $M(t) = \sum_{j \geq 0} 2^{-j} c_j(t)$, is also conserved, $M(t) = M(0) = 1$.

Equations (6.64) become linear if instead of t we use $\tau = \int_0^t dt' N(t') = -\ln(1 - t)$ as the time measure.[10] Solving $dc_n/d\tau = 2c_{n-1} - c_n$ subject to the monodisperse initial conditions $c_n(0) = \delta_{n,0}$, the density distribution is Poissonian: $c_n(\tau) = e^{-\tau} (2\tau)^n/n!$. In terms of the original time variable,

$$c_n(t) = (1 - t) \frac{[-2\ln(1 - t)]^n}{n!} . \tag{6.65}$$

At the shattering time $t_c = 1$ all densities vanish and the system undergoes a discontinuous (first-order) phase transition, all the entire mass is shattered into dust, and there are no particles with positive mass.

Example 6.5. *Larger particle splits*. Suppose now that the larger particle splits into two equal pieces; if the colliding particles have equal masses, a randomly chosen particle splits. The fragment mass density satisfies the master equation

$$\frac{d}{dt} c_n = 4c_{n-1}A_n - 2c_n A_{n+1} + 2c_{n-1}^2 - c_n^2, \tag{6.66}$$

where $A_n = \sum_{j \geq n} c_j$ is the cumulative density of fragments of mass 2^{-n} and smaller. It is again useful to verify that the mass density is conserved and that the fragment density satisfies (6.62).

The density $c_0(t)$ of unit-mass particles obeys $dc_0/dt = c_0^2 - 2c_0 N$. This equation is easy to solve because c_0^{-1} satisfies a linear equation. Using Eq. (6.63) and the initial condition $c_0(0) = 1$ we obtain

$$c_0(t) = \frac{3(1 - t)^2}{2 + (1 - t)^3} . $$

The fragment densities c_n with $n \geq 1$ satisfy Riccati equations, so there is no general technique for solving these equations. The asymptotic behavior of the density profile can nevertheless be extracted using a traveling wave technique (problem 6.13). Here we just seek the behavior near the shattering transition.

Since $A_n \to N$ as $t \to 1$, and the last two terms in (6.66) are asymptotically negligible, Eqs (6.66) simplify to $dc_n/d\tau = 2N(2c_{n-1} - c_n)$ which are identical (up to the factor of 2) to (6.64). Hence $c_n(\tau) \sim e^{-2\tau} (4\tau)^n/n!$, or

$$c_n \sim (1 - t)^2 \frac{[-4\ln(1 - t)]^n}{n!} . $$

---

[10] Physically, the modified time τ is the average number of collisions experienced by a fragment up to time t.

Apart from logarithmic corrections, the densities vanish quadratically: $c_n \sim (1-t)^2$. Thus the shattering transition is again discontinuous.

Our naive description of collisional fragmentation by bimolecular interactions has neglected the momenta of fragments, as well as the feature that fragment sizes are decreasing with time. We can rectify these shortcomings by a kinetic theory approach, parallel to that given in Section 3.7 to treat the increase in cluster radii during ballistic agglomeration. We start by rewriting the rate equation (6.62) as

$$\frac{dN}{dt} \sim \frac{N}{\tau}, \tag{6.67}$$

where τ is the average time between consecutive collision events. If all fragments have the same size independent of their mass, the time between collisions would scale as N^{-1}, thereby reproducing Eq. (6.62).

To determine τ when the size of fragments decreases, we employ the standard kinetic theory argument that the number of fragments in the collision tube that is swept out by the trajectory of a given fragment during a time τ should be of the order of one, namely, $N(V\tau)R^{d-1} \sim 1$. Since the average fragment mass scales as $m \sim N^{-1}$, the typical fragment radius scales as $R \sim m^{1/d} \sim N^{-1/d}$, which also gives the radius of the collision tube. Moreover, the typical fragment speed does not change with time due to energy conservation. Therefore

$$\tau \sim N^{-1} R^{-(d-1)} \sim N^{-1} N^{(d-1)/d} = N^{-1/d} .$$

Substituting this estimate for τ into (6.67), we obtain the rate equation

$$\frac{dN}{dt} \sim N^{1+1/d} . \tag{6.68}$$

This modified rate equation leads to a shattering transition in which

$$N \sim N_0 (1 - t/t_c)^{-d}, \qquad \text{with} \quad t_c \sim (V N_0 R_0^{d-1})^{-1} . \tag{6.69}$$

Note that the exponent is equal to the spatial dimension d; hence the simple result (6.63) is qualitatively correct only in one dimension.

We can use this more accurate framework to analyze the mass distribution. In the case where a random particle splits (as in Example 6.4), instead of Eqs (6.64) we get

$$\frac{dc_n}{dt} = 2c_{n-1} \sum_{j\geq 0} (R_{n-1} + R_j)^{d-1} c_j - c_n \sum_{j\geq 0} (R_n + R_j)^{d-1} c_j \tag{6.70}$$

with $R_n = 2^{-n/d}$. While these equations do not admit a linearization, we can extract the asymptotic behavior near the shattering transition. For example, the density of the largest

fragments can be found from

$$\frac{dc_0}{dt} = -c_0 \sum_{j \geq 0} (1 + R_j)^{d-1} c_j \sim -c_0 N$$

to give

$$\ln c_0 \sim -\frac{1}{(1 - t/t_c)^{d-1}} \qquad \text{when} \quad t \to t_c.$$

The densities of smaller fragments exhibit a similar extremely rapid vanishing with time as $t \to t_c$ when $d > 1$ so that the shattering transition remains discontinuous.

6.5 Notes

Early work on fragmentation is reviewed in [105]. A variety of exact solutions for fragmentation models with specific breakup rates have been found; a representative set of these publications are [106–110]. A scaling theory of fragmentation that mirrors the scaling theory for the complementary process of aggregation was developed in [111]. The shattering transition was first investigated by Filippov [112] and many more properties of this phenomenon were elucidated in [109, 111].

The solution for fragmentation with a steady input of material is given in [113]; it is based on the series expansion that was developed by Charlesby [114]. The rectangular fragmentation model was introduced in [115] and the methodology of the solution will also be used to treat the adsorption of needles in the next chapter. A standard reference on reversible polymerization is by Blatz and Tobolsky [116]. The ramifications of the detailed balance condition for reversible fragmentation were explored in [117]. The unusual kinetic features of the chipping model were first discussed in [118], while the non-equilibrium phase transition in this model was discovered in [119]. The collision-induced fragmentation model was introduced in [111] and many of its features are treated in [120].

6.6 Problems

6.1 Obtain the leading asymptotic behavior of the moments for the random scission model.

6.2 Solve the random scission model for an initial exponential length distribution and show that this distribution remains exponential.

6.3 Consider the homogeneous breakup kernel $R(x) = 1$, which has the homogeneity index $\lambda = 0$, and hence lies on the borderline between scaling and shattering.

(a) Show that for the monodisperse initial condition $c(x, t = 0) = \delta(x - \ell)$, the fragment mass distribution is

$$c(x, t) = e^{-t}\delta(x - \ell) + \frac{2t\, e^{-t}}{\ell} \sum_{n \geq 0} \frac{[2t \ln(\ell/x)]^n}{n!\,(n+1)!}.$$

(b) Express the series solution from (a) in terms of the modified Bessel function $I_1[2\sqrt{2t \ln(\ell/x)}]$.

(c) Show that the moments of the mass distribution are given by $M_n = \ell^n \exp[(1 - n)t/(1 + n)]$.

6.4 Obtain the steady-state solution of the master equation (6.1b) for arbitrary breakup rate $R(x) \neq 0$.

6.5 Consider the three-dimensional generalization of rectangular fragmentation from Section 6.2. Assume that the initial fragment is a $1 \times 1 \times 1$ cube and that each crack nucleates a planar cut in any of the three mutually orthogonal directions.
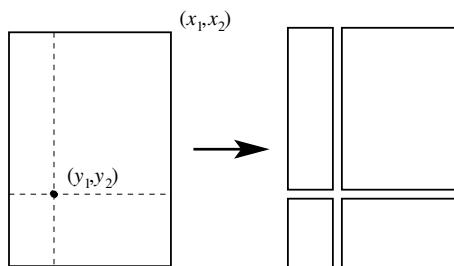
(a) Show that the Mellin transform satisfies

$$\frac{\partial c(s_1, s_2, s_3, t)}{\partial t} = \left[\frac{2}{3}\left(\frac{1}{s_1} + \frac{1}{s_2} + \frac{1}{s_3}\right) - 1\right]c(s_1 + 1, s_2 + 1, s_3 + 1, t).$$

(b) Seek the asymptotic behavior of the Mellin transform in the form $c(s_1, s_2, s_3, t) \sim t^{-\alpha(s_1, s_2, s_3)}$ and determine the exponent $\alpha(s_1, s_2, s_3)$.

(c) Verify that the volume distribution admits a scaling form $F(V, t) \simeq t^2 e^{-Vt}$, again the same as in the random scission model.

6.6 Consider the geometric fragmentation process in which a rectangle splits into four smaller fragments, as illustrated below.

(a) Verify that the Mellin transform $c(s_1, s_2, t)$ obeys

$$\frac{\partial c(s_1, s_2, t)}{\partial t} = \left(\frac{4}{s_1 s_2} - 1\right)c(s_1 + 1, s_2 + 1, t).$$

(b) Check that area conservation is satisfied by the above master equation and show that the average number of fragments is equal to $1 + 3t$.

(c) Using the equation for $c(s_1, s_2, t)$ from (a) and the asymptotic behavior $c(s_1, s_2, t) \sim t^{-\alpha(s_1, s_2)}$, show that

$$\alpha(s_1, s_2) = \left[(s_1 + s_2) - \sqrt{(s_1 - s_2)^2 + 16}\right]/2.$$

(d) Show that the area distribution admits a scaling form

$$F(A, t) \simeq t^{-2} f(At), \qquad f(z) = 6 \int_0^1 d\xi \, (\xi^{-1} - 1) e^{-z/\xi}.$$

(e) Using the results from (c) show that the area distribution weakly diverges in the limit of small area, $f(z) \simeq 6 \ln(1/z)$. Verify that in the opposite limit the scaled area distribution exhibits an exponential decay reminiscent of the one-dimensional random scission model, but with an algebraic correction: $f(z) \simeq 6z^{-2} \exp(-z)$ as $z \gg 1$.

6.7 Consider a three-state system with non-zero transition rates $w_{i \to j} \neq 0$ for all $i \neq j$. What condition or conditions must these rates meet for the steady state to satisfy detailed balance (6.45)?

6.8 Consider combined polymerization/fragmentation in the continuum limit, where the polymer length is a continuous variable.

(a) Show that the master equation for this continuum model is

$$\frac{\partial c(\ell)}{\partial t} = \int_0^\ell c(x) \, c(\ell - x) \, dx - 2c(\ell) \int_0^\infty c(x) \, dx + \lambda \left[2 \int_\ell^\infty c(x) \, dx - \ell \, c(\ell) \right].$$

(v) Verify that the stationary solution to this master equation is given by $c(\ell) = \lambda e^{-\ell/\ell_0}$ and determine ℓ_0.

6.9 Consider the evolution of closed strings at high temperature. The goal is to generalize the master equation for open strings in the previous problems to the case of closed strings and to determine the stationary solution.

(a) Argue that the master equation is

$$\frac{\partial c(\ell)}{\partial t} = \int_0^\ell x(\ell - x) \, c(x) \, c(\ell - x) \, dx - 2\ell \, c(\ell) \int_0^\infty x \, c(x) \, dx$$
$$+ \lambda \left[2 \int_\ell^\infty x \, c(x) \, dx - \ell^2 \, c(\ell) \right].$$

The first two terms on the right-hand side describe aggregation events – in contrast to open strings where the joining occurs when the ends meet, for closed strings every intersection of two loops can result in aggregation. The last two terms describe fragmentation events that occur when a closed string self-intersects.

(b) Show that the above master equation admits the stationary solution $c = (\lambda/\ell) e^{-\ell/\ell_0}$.

(c) Show that the same length distribution follows from the detailed balance condition.

6.10 Show that for the chipping model the fragment density is equal to $N = \lambda/2$ in the gel phase ($\lambda < 1$), thus establishing the second line of (6.56).

6.11 Verify the expressions (6.59) and (6.61) for the mass density of the chipping model in the sol and gel phases.

6.12 Consider the collisional fragmentation model from Example 6.4, where one of the two colliding particles splits in two.

(a) Show that near the shattering transition, the Poisson form (6.65) for mass density approaches a Gaussian distribution.

(b) Show that the typical mass m_* (corresponding to the peak of the mass distribution) decays to zero as

$$m_* \sim (1-t)^\sigma \qquad \text{with} \quad \sigma = 2\ln 2 \doteq 1.386294 \,.$$

6.13 Consider the collisional fragmentation model from Example 6.5, where the larger of the two colliding particles splits in two.

(a) Show that in terms of the cumulative densities, Eq. (6.66) simplifies to

$$\frac{dA_n}{dt} = 2A_{n-1}^2 - A_n^2 \,.$$

(b) Normalizing the cumulative distribution, $F_n(\tau) = N^{-1}A_n(t)$, recast the master equation from (a) into

$$\frac{dF_n}{d\tau} = 2F_{n-1}^2 - F_n^2 - F_n \,.$$

(c) Argue that the master equation for F admits a traveling wave solution $F_n(\tau) \to f(n - v\tau)$. Substitute this wave form into the above equation for F to deduce

$$v\frac{d}{dx}f(x) = f(x) + f^2(x) - 2f^2(x-1) \,.$$

(d) Show that the boundary conditions for the above equation are $f(-\infty) = 1$ and $f(\infty) = 0$. Try to guess the approach to final asymptotics in the $x \to \pm\infty$ limits.

(e) To determine the speed v, seek the asymptotic behavior of $f(x)$ far behind the front in the form $1 - f(x) \sim e^{\lambda x}$ as $x \to -\infty$. Substitute this form into the equation given in part (c) to deduce the "dispersion" relation $v = (3 - 4e^{-\lambda})/\lambda$ between the speed v and the decay coefficient λ. Plot $v = v(\lambda)$ for $\lambda > 0$ and show that the spectrum of possible velocities is bounded from above: $v \in (-\infty, v_{\max}]$.

(f) Out of the spectra of possible velocities $v \in (-\infty, v_{\max}]$, the maximal value is selected. Show that $v = v_{\max} \doteq 1.52961$.

(g) Show that the typical mass decays to zero according to Eq. (12) with $\sigma = v\ln 2 \doteq 1.06024$.

6.14 Consider a model where a randomly chosen collision partner splits stochastically, namely a particle of masses m splits into two fragments of mass m' and $m - m'$ with m' chosen uniformly in $[0, m]$.

(a) Show that the mass density $c(m, \tau)$ satisfies

$$\frac{\partial}{\partial\tau}c(m, \tau) = -c(m, \tau) + 2\int_m^\infty \frac{dm'}{m'}c(m', \tau) \,.$$

(b) Solve the master equation from (a) using the Mellin transform. Show that for the initial condition $c(m, 0) = \delta(m - 1)$, the mass distribution is

$$c(m, \tau) = e^{-\tau} \delta(m-1) + e^{-\tau} \sqrt{\frac{2\tau}{\ln(1/m)}} \, I_1 \left[\sqrt{8\tau \ln(1/m)} \right]$$

with I_1 the modified Bessel function.

(c) Show that, apart from the logarithmic factor, $c(m, t) \sim 1 - t$ near the shattering transition. (Essentially the same behavior was found for collisional fragmentation with a deterministic splitting rule.)
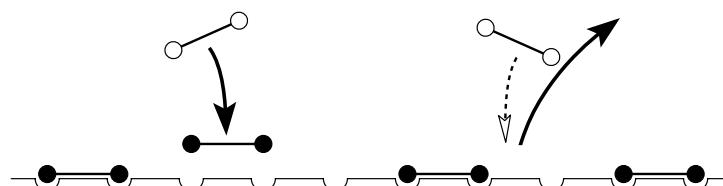
Adsorption

Suppose that gas molecules impinge upon and adsorb on a surface, or substrate. If the incident molecules are monomers that permanently attach to single adsorption sites on the surface and there are no interactions between adsorbed monomers, then the density ρ of occupied sites increases with time at a rate proportional to the density of vacancies, namely $d\rho/dt = 1 - \rho$. Thus $\rho(t) = 1 - e^{-t}$, and vacancies disappear exponentially in time. However, if each arriving molecule covers more than one site on the substrate, then a vacant region that is smaller than the molecular size can never be filled. The substrate reaches an incompletely filled *jammed* state that cannot accommodate additional adsorption. What is the filling fraction of this jammed state? What is the rate at which the final fraction is reached? These are basic questions of adsorption kinetics.

7.1 Random sequential adsorption in one dimension

Dimer adsorption

A simple example with non-trivial collective behavior is the *random sequential adsorption* of dimers – molecules that occupy two adjacent sites of a one-dimensional lattice (Fig. 7.1). We model the steady influx of molecules by adsorption attempts occurring one at a time at random locations on the substrate. An adsorption attempt is successful only if a dimer lands on two adjacent empty sites. If a dimer lands on either two occupied sites or on one occupied and one empty site, the attempt fails. After each successful attempt, the coverage increases. Eventually only isolated vacancies remain, and whenever this state is reached, the substrate is jammed since no further adsorption is possible.

For dimer adsorption, a jammed state consists of strings with an even number of occupied sites (●) that are separated by isolated vacant sites (○), as illustrated in Fig. 7.2. In

Fig. 7.1. Illustration of irreversible dimer deposition. The dimer on the left successfully adsorbs onto two adjacent vacant sites, while the dimer on the right does not adsorb.

Fig. 7.2.   A jammed configuration in random sequential adsorption of dimers.

principle, the fraction of occupied sites in the jammed state, $\rho_{\text{jam}} \equiv \rho(t = \infty)$, can have any value between 2/3 and 1, with the two extreme limits achieved by the respective configurations of one empty site between adsorbed dimers ($\cdots \bullet\!\bullet \circ \bullet\!\bullet \circ \bullet\!\bullet \cdots$) or no empty sites ($\cdots \bullet\!\bullet \; \bullet\!\bullet \; \bullet\!\bullet \cdots$). A beautiful result, first derived by Flory using a combinatorial approach, is that an initially empty substrate saturates to a jamming density, ρ_{jam}, whose value is

$$\rho_{\text{jam}} = 1 - e^{-2} \doteq 0.864664. \tag{7.1}$$

Empty interval method

Here we adopt the kinetic viewpoint and determine the complete time evolution of the coverage. The final coverage then emerges as a simple consequence. The full description of the evolution requires knowing the probabilities of all possible configurations of adsorbed molecules on the substrate. However, these probabilities contain more information than necessary for determining the evolution of the coverage. What we really need in one dimension are the *empty interval probabilities*, namely, the probability that a string of *m consecutive* sites are empty:

$$E_m \equiv \mathcal{P}[\underbrace{\circ \cdots \circ}_{m}]. \tag{7.2}$$

Here $\mathcal{P}$ denotes the probability of a configuration. Notice that the states of the sites external to the string remain unspecified. Consequently, E_m is the probability to find an empty interval of length *m or greater*. In particular, E_1 is the density of empty sites and $\rho = 1 - E_1$ is the density of occupied sites.

For irreversible dimer adsorption, the equations that describe the evolution of the empty interval probabilities E_m are

$$\frac{dE_m}{dt} = -(m - 1)E_m - 2E_{m+1}, \qquad m \geq 1. \tag{7.3}$$

We typically set the intrinsic adsorption rate to 1, corresponding to measuring time in units of the inverse adsorption rate. The first term on the right-hand side of Eq. (7.3) accounts for the disappearance of an *m*-interval due to the adsorption of dimers inside the interval. The factor $m - 1$ counts the $m - 1$ distinct locations where the dimer can adsorb, such that it lies entirely within the interval (Fig. 7.3). The second term accounts for the two distinct adsorption events in which one end of the incident dimer is outside the *m*-interval. For these latter events, the empty interval must contain at least $m + 1$ empty sites to accommodate adsorption of a dimer on the periphery of the *m*-interval; hence the factor E_{m+1}. Notice that Eq. (7.3) contains only loss terms; this feature is a consequence of using the empty interval probabilities E_m as the basic dynamical variables. Effectively, we are looking at
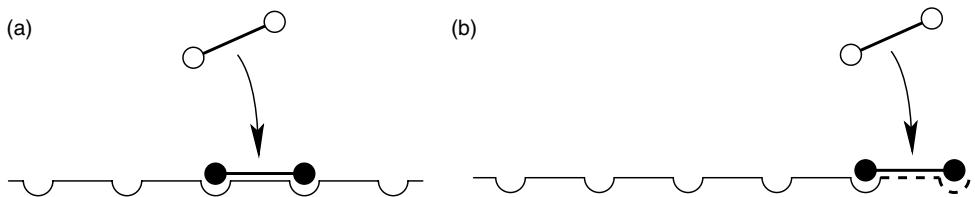
Fig. 7.3.   Changes in the empty interval probability E_m for $m = 5$, with an adsorption event (a) in the interior of the interval and (b) at the edge of the interval (dashed portion).

the substrate through lenses that see only m consecutive sites at a time and E_m is merely the fraction of these intervals that are empty. In this representation, there is no way to create an empty interval of length $\geq m$ by the adsorption of a dimer onto a still-larger empty interval.

To solve (7.3), notice that the homogeneous part of this equation admits the solution $e^{-(m-1)t}$. This fact suggests seeking a solution to the full equation that is of the form

$$E_m(t) = e^{-(m-1)t} \, \Phi(t), \tag{7.4}$$

where $\Phi(t)$ coincides with $E_1(t)$. For the natural initial condition of an empty system, $E_m(0) = 1$, which gives $\Phi(0) = 1$. Using this ansatz for $E_m(t)$, the infinite set of evolution equations (7.3) miraculously reduces to the *single* equation $d\Phi/dt = -2e^{-t} \, \Phi$, with solution $\Phi(t) = \exp\left[-2(1 - e^{-t})\right]$. Then the empty interval probabilities are

$$E_m(t) = \exp\left[-(m-1)t - 2\left(1 - e^{-t}\right)\right]. \tag{7.5}$$

Empty gaps of length greater than 1 decay exponentially with time and only gaps of length 1 remain in the final jammed state. From (7.5), the asymptotic density of such gaps is $E_1(\infty) = e^{-2}$, so that the jamming coverage is $\rho_{\text{jam}} = 1 - e^{-2}$. We have therefore reproduced Flory's result with little labor, but we also have derived much more – the coverage throughout the entire evolution:

$$\rho(t) = 1 - \exp\left[-2\left(1 - e^{-t}\right)\right]. \tag{7.6}$$

The coverage approaches the jamming value exponentially in time, $\rho_{\text{jam}} - \rho(t) \simeq C\,e^{-t}$, with $C = 2\,e^{-2}$; this rapid decay to the final state typifies lattice models of irreversible adsorption (problems 7.3 and 7.4).

Empty intervals and voids

An alternative to the empty interval probabilities E_m as the basic variables for irreversible adsorption are the *void probabilities*. A void of length m is defined as a group of m consecutive empty sites that are bounded on both sides by an occupied site. Pictorially, the probability for an m-void is defined by

$$V_m \equiv \mathcal{P}[\bullet \underbrace{\circ \cdots \circ}_{m} \bullet].$$

To relate the void probabilities V_m to the empty interval probabilities E_m, let's start with the simplest configurations. Clearly, $\mathcal{P}[\bullet] + \mathcal{P}[\circ] = \rho + E_1 = 1$. For two sites, the following conservation statements hold:

$$\mathcal{P}[\circ\circ] + \mathcal{P}[\circ\bullet] = \mathcal{P}[\circ] \quad \text{and} \quad \mathcal{P}[\circ\circ] + \mathcal{P}[\circ\bullet] + \mathcal{P}[\bullet\circ] + \mathcal{P}[\bullet\bullet] = 1 \,;$$

while for intervals of arbitrary length:

$$\mathcal{P}[\underbrace{\circ\cdots\circ}_{m}\bullet] + \mathcal{P}[\underbrace{\circ\cdots\circ\circ}_{m+1}] = \mathcal{P}[\underbrace{\circ\cdots\circ}_{m}] \,. \tag{7.7a}$$

We also define S_m, the auxiliary *segment probability*, in which the state of the site at one end of the segment is specified, as

$$S_m \equiv \mathcal{P}[\underbrace{\circ\cdots\circ}_{m}\bullet] \,. \tag{7.7b}$$

Using (7.7a), the segment probability may be rewritten in terms of empty interval probabilities, $S_m = E_m - E_{m+1}$. Similarly, the void probability is a difference of segment probabilities,

$$V_m = S_m - S_{m+1} = E_m - 2E_{m+1} + E_{m+2} \,. \tag{7.8a}$$

In the continuum limit, the correspondences between the void density $V(x, t)$, the segment density $S(x, t)$, and the empty interval density $E(x, t)$ become

$$S(x, t) = -\frac{\partial E(x, t)}{\partial x} \quad \text{and} \quad V(x, t) = -\frac{\partial S(x, t)}{\partial x} = \frac{\partial^2 E(x, t)}{\partial x^2} \,. \tag{7.8b}$$

There are no hard-and-fast rules for which set of quantities – E_m or V_m – are more useful for determining the dynamics of adsorption and other kinetic problems. While the void probabilities may be conceptually simpler at first sight, the evolution equations for V_m contain both gain and loss terms (problem 7.5) and can be more complicated to analyze compared to the corresponding equations for E_m.

Simulating adsorption

Without loss of generality, we set the adsorption rate to one. There are two basic methods to simulate dimer adsorption in one dimension.

- *Dumb algorithm*. At each simulation step: (i) pick one site at random; (ii) accept the adsorption attempt at this site if and only if the site and its neighbor to the right are vacant; and (iii) update the time by the inverse of L, the total number of sites: $t \rightarrow t + L^{-1}$. Repeat this simulation step until a prescribed coverage is reached. This algorithm is easy to implement but is computationally wasteful because the vast majority of adsorption attempts fail at late times.

- *Smart algorithm*. At each simulation step: (i) pick one *allowed* site at random; (ii) adsorb a dimer at this allowed site and its right neighbor; and (iii) update the time by an amount
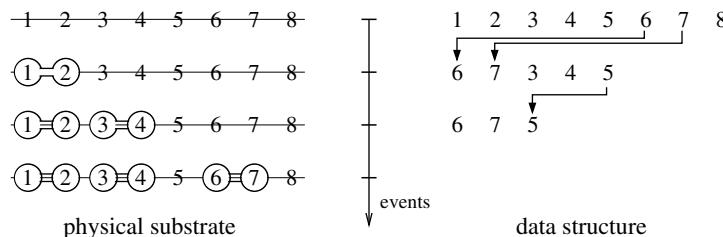
Fig. 7.4.

Efficient simulation of dimer adsorption on a ring of $L = 8$ sites. The left side shows the physical substrate after each individual adsorption event and the right side shows the corresponding list of allowed adsorption sites (periodic boundary condition assumed). This numerical list is compacted at each step by replacing the locations of no-longer-allowed sites with locations of allowed sites. For example, after the first adsorption event, sites 1, 2, and 8 are no longer allowed. We relabel so that the first allowed site is now 6, the second allowed site is 7, while site 8 is simply removed.

inversely proportional to L_a, the total number of allowed adsorption sites: $t \rightarrow t + L_a^{-1}$. Repeat this simulation step until jamming is reached ($L_a = 0$). This algorithm requires us to keep track of all allowed adsorption sites. This additional overhead in bookkeeping (Fig. 7.4) pays off in computational efficiency because each adsorption attempt is always successful, and there is a fixed computational cost per adsorption event. Figure 7.4 illustrates this bookkeeping.

In general, this algorithm can be implemented to simulate the adsorption of arbitrary shapes on both discrete and continuum substrates by continuously keeping track of the allowed space for further adsorption.

Adsorption of longer molecules

What if the incident molecules occupy k consecutive substrate sites? (Such linear polymers are called k-mers.) The coverage for this general adsorption process can be obtained by a straightforward extension of our approach for dimers. Now the equations for the empty interval probabilities E_m in k-mer adsorption are

$$\frac{dE_m}{dt} = \begin{cases} -(m-k+1)E_m(t) - 2 \displaystyle\sum_{1 \leq j \leq k-1} E_{m+j}(t), & m \geq k, \\[2ex] -(k-m+1)E_k(t) - 2 \displaystyle\sum_{1 \leq j \leq m-1} E_{k+j}(t), & m \leq k. \end{cases} \tag{7.9}$$

The terms in this equation mirror those in Eqs (7.3) for dimer adsorption. For $m \geq k$, the first term accounts for the $m - k + 1$ distinct ways that a k-mer can adsorb in the interior of an empty interval of size m. The second term accounts for the $2(k-1)$ ways that the k-mer can adsorb with the k-mer partially outside and partially inside the original m-interval. For $m < k$, the first term accounts for the $k - m + 1$ ways that the m-interval can be completely covered by the k-mer. The second term accounts for the ways in which

the k-mer partially covers the interval. Notice that the equation for $m < k$ can be obtained merely by interchanging the roles of k and m in the equation for $m \geq k$.

In analogy with Eq. (7.4), we make the ansatz $E_m(t) = e^{-(m-k+1)t} \Phi(t)$, substitute into Eq. (7.9) for $m \geq k$, to yield $d\Phi/dt = -2\Phi \sum_{1 \leq j \leq k} e^{-jt}$. Solving, we find Φ, from which the empty interval probabilities are

$$E_m(t) = \exp\left[-(m-k+1)t - 2 \sum_{1 \leq j \leq k-1} \frac{1-e^{-jt}}{j}\right], \qquad m \geq k. \qquad (7.10)$$

To find the time dependence of the coverage, $\rho = 1 - E_1$, we consider the second of Eqs (7.9) for the case $m = 1$. This gives $dE_1/dt = -k E_k$, or equivalently $d\rho/dt = k E_k$. However, the quantity E_k is already known from Eq. (7.10) with $m = k$. Integrating the latter, we obtain the coverage

$$\rho(t) = k \int_0^t \exp\left[-u - 2 \sum_{1 \leq j \leq k-1} \frac{1-e^{-ju}}{j}\right] du. \qquad (7.11)$$

Numerical evaluation of this integral gives a jamming coverage that decreases monotonically with k; for example, $\rho_{\text{jam}} \doteq 0.864665, 0.823653, 0.803893$, and 0.792276 for $k = 2$–5 (Fig. 7.5). As the molecule size increases, the coverage of the jammed state decreases because progressively longer empty regions can exist.

Irreversible car parking

The limit of k-mer adsorption with $k \to \infty$ defines the *car parking* problem. In this limit, the position of an adsorbed k-mer becomes continuous and one can think of unit-length
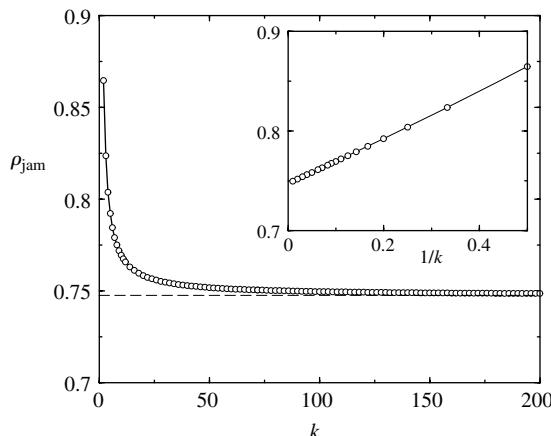
Fig. 7.5.   Dependence of ρ_{jam} on the length k of the adsorbed polymer. The dashed line is the asymptote $\rho_{\text{jam}} \doteq 0.747597$. The inset shows the same data plotted versus $1/k$.

"cars" that irreversibly park anywhere along a one-dimensional curb (no marked parking spots) and are then abandoned. The only constraint is that cars cannot overlap; however, a car *can* fit into a parking spot that is infinitesimally larger than the car itself. For this parking problem, Rényi obtained a jamming coverage of $\rho_{\mathrm{jam}} \doteq 0.747597$.

We again solve the car parking problem by a master equation approach. The length of a car is immaterial if we seek the fraction of the line that is covered, and therefore we set the car length to be 1. The appropriate dynamical variable is now $E(x, t)$, the probability that a randomly chosen interval of length x is empty. As in the discrete case, such a region may be part of an even longer empty interval. When the rate at which cars park equals 1, the evolution of $E(x, t)$ is described by (compare with Eqs (7.9))

$$\frac{\partial E(x, t)}{\partial t} = \begin{cases} -(x - 1)E(x, t) - 2 \displaystyle\int_x^{x+1} E(y, t) \, dy, & x > 1, \\[4mm] -(1 - x)E(1, t) - 2 \displaystyle\int_1^{x+1} E(y, t) \, dy, & x < 1. \end{cases} \tag{7.12}$$

The terms in this equation have direct counterparts with those in (7.9) for k-mer adsorption. For $x > 1$, the first term on the right in Eq. (7.12) accounts for adsorption events that lie completely within the interval, while the second term accounts for adsorption events that partially overlap the interval. A similar correspondence also applies for the second line of (7.12), and again notice the interchange of the role of the variables x and 1.

It is also instructive to write the master equations for the void density $V(x, t)$,

$$\frac{\partial V(x, t)}{\partial t} = \begin{cases} -(x - 1)V(x, t) + 2 \displaystyle\int_{x+1}^{\infty} V(y, t) \, dy, & x > 1, \\[4mm] 2 \displaystyle\int_{x+1}^{\infty} V(y, t) \, dy, & x < 1, \end{cases} \tag{7.13}$$

to appreciate the correspondence with the empty interval probabilities. For $x > 1$ the first term accounts for adsorption in the interior of an interval of length x, exactly as in Eq. (7.12). The second term accounts for the creation of an x-void by the "fragmentation" of an $(x+1)$-void by adsorption inside this larger void (note the close analogy with the master equation for fragmentation, Eq. (6.2)). There are exactly two locations for this adsorption event that give rise to an x-void. For $x < 1$, there is no loss term because adsorption cannot occur inside such a short void, but there is still a gain term due to the fragmentation of a void of length $x + 1$ or greater.

To solve for $E(x, t)$, consider first the regime $x > 1$. As in the discrete case, we seek a solution of the form $E(x, t) = e^{-(x-1)t} E(1, t)$. Substituting this expression into the first of Eqs (7.12), the x-dependent terms cancel, and integrating the resulting equation for $E(1, t)$ gives

$$E(1, t) = \exp\left[-2 \int_0^t \frac{1 - e^{-u}}{u} \, du \right], \tag{7.14}$$

from which we immediately obtain $E(x, t)$ for $x > 1$. From the second of (7.12), we have $\partial E(0, t)/\partial t = -E(1, t)$, so that the coverage $\rho(t) = 1 - E(0, t)$ is

$$\rho(t) = \int_0^t \exp\left[-2\int_0^v du \, \frac{1 - e^{-u}}{u}\right] dv. \tag{7.15}$$

Numerical evaluation of this integral for $t \to \infty$ gives the jamming coverage $\rho_{\text{jam}} \doteq 0.747597$.[1]

A qualitatively new feature of continuum car parking is that the approach to jamming is considerably slower than that for adsorption onto a lattice. Let's examine how the deviation from the asymptotic coverage, $\rho_{\text{jam}} - \rho(t)$, vanishes as $t \to \infty$. For this purpose, we write the asymptotic behavior of the integral in Eq. (7.15) as

$$\int_0^t \frac{1 - e^{-u}}{u} \, du = \ln t + \gamma_E + t^{-1} e^{-t} + \cdots,$$

where $\gamma_E \doteq 0.577215$ is the Euler constant. With this result, the approach to the jamming coverage is given by

$$\begin{aligned}
\rho_{\text{jam}} - \rho(t) &= \int_t^\infty \exp\left[-2\int_0^v \frac{du}{u}(1 - e^{-u})\right] dv \\
&\simeq \int_t^\infty \exp\left[-2(\ln v + \gamma_E)\right] dv \\
&= e^{-2\gamma_E} \, t^{-1}.
\end{aligned} \tag{7.16}$$

This slow approach to jamming arises because parking spaces that are infinitesimally longer than a car can arise and such spaces are filled extremely slowly. In d spatial dimensions, a related power-law relaxation occurs in which the density relaxes to the jamming density as $t^{-1/d}$ (see Section 7.3).

7.2 Phase space structure

Broken ergodicity

A basic tenet of equilibrium statistical physics is the postulate of *equal a priori probability* – each microscopic state with the same energy is realized with the same probability. This feature is equivalent to the notion of *ergodicity* that underlies the equivalence between the formal ensemble average of statistical mechanics and the time average in a real equilibrium system. Non-equilibrium systems in general and systems with jammed configurations in particular do not fit into this picture as they do not uniformly sample the state space of all configurations. This phenomenon is known as *broken ergodicity*. Irreversible adsorption is an ideal setting to understand broken ergodicity because all states can be readily computed.

---

[1] This same result can also be obtained by taking the limit of Eq. (7.11) with $k \to \infty$ and $t \to \infty$.

Let's start by counting the total number of jammed states in a finite system for irreversible dimer adsorption. Let J_L be the total number of jammed configurations on a finite chain of L sites. We can readily count the number of jammed configurations on small chains: $J_2 = 1$ (●●), $J_3 = 2$ (●● ○ and ○ ●●), $J_4 = 2$ (●● ●● and ○ ●● ○), and $J_5 = 3$ (●● ●● ○, ○ ●● ●●, and ●● ○ ●●). In general, there are two types of jammed configurations: (i) those in which the first site of the chain is occupied, and (ii) those in which the first site is empty. Typical configurations in these two categories look like:

$$\underbrace{\bullet\!\bullet \; \times \cdots \times}_{L-2} \qquad \text{and} \qquad \circ \; \bullet\!\bullet \; \underbrace{\times \cdots \times}_{L-3} ,$$

respectively. The first category therefore consists of J_{L-2} configurations and the second category consists of J_{L-3} configurations. Thus J_L satisfies the Fibonacci-like recurrence

$$J_L = J_{L-2} + J_{L-3} \tag{7.17}$$

for $L > 2$, subject to the boundary conditions $J_0 = J_1 = J_2 = 1$. For the asymptotic solution, we substitute the exponential form $J_L = z^L$ into the recursion for J_L and see that z must satisfy the cubic polynomial $z^3 - z - 1 = 0$. Only one of these roots ($z \doteq 1.32472$) is real and has magnitude greater than 1. As a result, the asymptotic behavior of the number of jammed configurations does indeed grow exponentially with system size (see problem 7.8),

$$J_L \sim z^L , \tag{7.18}$$
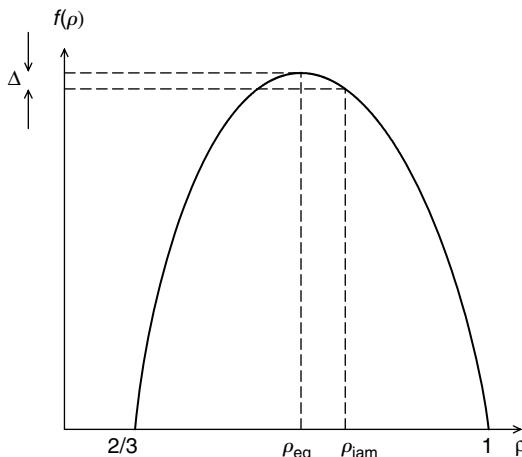
so that the entropy of packing is $S = \ln J_L \simeq L \ln z$.

It is now instructive to determine the number of jammed configurations with a *specified* coverage. Let $J_{N,L}$ be the number of jammed configurations that contain N dimers in a system of size L. The number of dimers must be in the range $\lfloor (L+1)/3 \rfloor \leq N \leq \lfloor L/2 \rfloor$ for the configuration to actually be jammed, with $\lfloor x \rfloor$ the integer part of x. In a jammed configuration, a dimer must be followed either by another dimer or by a *single* vacant site. Thus each distinct jammed configuration may be symbolically written as $\ldots$DDODDOD$\ldots$. That is, between each pair of dimers there may be either a single vacancy or nothing. Since a vacancy can appear between any pair of dimers and also between a dimer and the end of the chain, there are $N + 1$ possible locations for the $L - 2N$ vacant sites. Therefore the number of distinct arrangements with N dimers is given by the binomial coefficient

$$J_{N,L} = \binom{N+1}{L-2N}, \tag{7.19}$$

and the total number of allowable configurations is $J_L = \sum_N J_{N,L}$.

In the thermodynamic limit, we fix the coverage to $\rho = 2N/L$ and evaluate the total number of configurations with a given density, $J_L(\rho) \equiv J_{\rho L/2, L}$, by keeping only the two leading terms in the Stirling formula $\ln x! \sim x \ln x - x$. In this approximation, the total number of

Fig. 7.6. The entropy function. Indicated are the jamming and equilibrium densities, and the entropy difference Δ.

fixed-density configurations grows as $J_L(\rho) \sim e^{Lf(\rho)}$, with entropy function (Fig. 7.6)

$$f(\rho) = \frac{\rho}{2} \ln \frac{\rho}{2} - (1 - \rho) \ln(1 - \rho) - \left(\frac{3\rho}{2} - 1\right) \ln \left(\frac{3\rho}{2} - 1\right) . \qquad (7.20)$$

There are several important features of $f(\rho)$. First, the total number of jammed configurations is exponentially large for any density $2/3 < \rho < 1$ because $f(\rho)$ is positive in this range.[2] Second, because $J_L(\rho)$ grows exponentially with L, $J_L(\rho)$ is dominated by its most probable value. The condition $f'(\rho) = 0$ gives $\frac{1}{2}\rho(1 - \rho)^2 = \left(\frac{3}{2}\rho - 1\right)^3$, from which the equilibrium density is $\rho_{\mathrm{eq}} \doteq 0.822991$. Close to the maximum, $f(\rho) = f(\rho_{\mathrm{eq}}) + \frac{1}{2}f''(\rho_{\mathrm{eq}})(\rho - \rho_{\mathrm{eq}})^2 + \cdots$, from which the density dependence of the number of jammed configurations approaches the Gaussian

$$J_L(\rho) \simeq \frac{J_L}{\sqrt{2\pi\sigma^2}} \exp\left[-(\rho - \rho_{\mathrm{eq}})^2/2\sigma^2\right] , \qquad (7.21)$$

with variance $\sigma^2 = [-f''(\rho_{\mathrm{eq}})L]^{-1}$.

We use the term "equilibrium" for ρ_{eq} since this density arises when all jammed states are sampled equiprobably. In fact, the Gaussian distribution (7.21) becomes sharply peaked so that the coverage in a randomly chosen jammed configuration almost surely matches the equilibrium density ρ_{eq}. For irreversible adsorption, however, the system explores a tiny part of phase space with a higher coverage, $\rho_{\mathrm{jam}} \doteq 0.864644 > \rho_{\mathrm{eq}} \doteq 0.822991$. The relative weight of this phase space region decays exponentially with system size, $e^{-\Delta L}$, with $\Delta = f(\rho_{\mathrm{eq}}) - f(\rho_{\mathrm{jam}})$ (Fig. 7.6). For non-equilibrium systems, as exemplified here by irreversible adsorption, the dynamics dictates how phase space is explored, and there is no reason for all microscopic configurations to be sampled uniformly.

---

[2] The number of configurations is finite at the two extremes: $J_L(2/3) = 2$ and $J_L(1) = 1$ and hence $f(2/3) = f(1) = 0$.

Finite-size corrections

The master equation analysis of adsorption kinetics describes only infinitely large systems, while real systems are finite. To appreciate the role of finiteness, we study irreversible dimer adsorption on a finite interval of length L by adopting Flory's combinatorial approach. Let A_L be the average number of occupied sites in all jammed states for dimer adsorption on an interval of length L. We can directly evaluate this quantity for small systems.

Example 7.1. *Jamming in small systems.* For a system that consists of a single site, A_1 clearly equals zero. For a two-site or a three-site interval, only a single dimer can adsorb, so that $A_2 = A_3 = 2$. The first non-trivial case is a four-site interval. With probability $\frac{1}{3}$ the first dimer lands on the middle two sites and no further adsorption is possible. With probability $\frac{2}{3}$, the first dimer occupies one of the interval endpoints and its neighbor so that a second dimer can be accommodated. Thus the average number of occupied sites is $\frac{1}{3} \times 2 + \frac{2}{3} \times 4 = \frac{10}{3}$. In summary, $A_1 = 0, A_2 = A_3 = 2, A_4 = \frac{10}{3}$, etc.

It is impracticable to carry out this step-by-step enumeration by hand. However, we can determine A_L for general L, by writing a recursion for the jamming density on a finite interval of length L in terms of the jamming densities on shorter intervals. If the first dimer lands on sites $(j, j+1)$ with $1 \leq j \leq L-1$, then two intervals of lengths $j-1$ and $L-j-1$ are created that subsequently get filled independently. Therefore, for $L \geq 2$, the coverage obeys

$$A_L = \frac{1}{L-1} \sum_{j=1}^{L-1} \left(2 + A_{j-1} + A_{L-j-1}\right) = 2 + \frac{2}{L-1} \sum_{j=0}^{L-2} A_j, \qquad (7.22)$$

with the boundary condition $A_0 = A_1 = 0$. That is, the number of occupied sites in the jammed state equals the sum of 2, for the initial dimer, plus the number of sites occupied in the two remaining sub-intervals. We combine these two contributions because, by symmetry, they are identical.

To solve for A_L, we use the generating function $\mathcal{A}(z) = \sum_{L \geq 2} A_L z^{L-1}$. Multiplying Eq. (7.22) by $(L-1) z^{L-2}$ and summing over all $L \geq 2$, we transform the left-hand side into $d\mathcal{A}/dz$ and the first term on the right-hand side into $2/(1-z)^2$. For the second term on the right-hand side, we interchange the order of summations to give

$$2 \sum_{L \geq 2} z^{L-2} \sum_{j=0}^{L-2} A_j = 2 \sum_{j \geq 0} A_j \sum_{L \geq j+2} z^{L-2} = 2 \sum_{j \geq 0} A_j \frac{z^j}{1-z} = \frac{2z}{1-z} \mathcal{A}(z).$$

With these steps we recast the recurrence (7.22) into the differential equation

$$\frac{d\mathcal{A}}{dz} = \frac{2}{(1-z)^2} + \frac{2z}{1-z} \mathcal{A}. \qquad (7.23)$$

The boundary conditions are $\mathcal{A}(z = 0) = A_1 = 0$ and $\mathcal{A}'(z = 0) = A_2 = 2$. To solve Eq. (7.23), notice that the integrating factor $I = (1-z)^2 e^{2z}$ reduces Eq. (7.23) to $dI/dz =$

$2\,e^{2z}$. The solution to Eq. (7.23) is

$$\mathcal{A} = \frac{1 - e^{-2z}}{(1 - z)^2}.$$ (7.24)

From the small-z behavior, $\mathcal{A} = 2z + 2z^2 + \frac{10}{3}z^3 + \cdots$, we rederive the average number of occupied sites in the jammed state for small systems.

As $L \to \infty$, we can obtain the series coefficients in the generating function from the behavior of $\mathcal{A}(z)$ as $z \uparrow 1$. If the coverage is linear in the system size, that is, $A_L \to L\rho + C$ as $L \to \infty$, with ρ and C constants, then in this limit $\mathcal{A}(z)$ must have the form

$$\mathcal{A}(z) = \sum_{L \geq 2} \left(\rho L + C\right) z^{L-1} = \rho \frac{d}{dz} \sum_{L \geq 2} z^L + C \sum_{L \geq 2} z^{L-1} = \rho \frac{z(2-z)}{(1-z)^2} + C \frac{z}{1-z}.$$ (7.25)

Comparing Eqs (7.24) and (7.25) as $z \to 1$, we find $\rho = \rho_{\text{jam}} = 1 - e^{-2}$, thereby recovering the Flory result (7.1). We also find the leading asymptotic correction, $C = -2e^{-2}$. Therefore the coverage in a system of size L, $\rho(L) \equiv A_L/L$, also contains a finite-size correction of the form

$$\rho(L) \simeq \rho_{\text{jam}} + C\,L^{-1}.$$ (7.26)

Such corrections to the asymptotic behavior due to finiteness often have a power-law form, as in (7.26). These finite-size corrections are useful in numerical calculations because these allow us to extrapolate infinite-system properties from finite-system simulations (problem 7.9). The magnitude of the correction term (proportional to L^{-1}) stems from the density being close to the Flory value except for a few sites at the edges of the interval.

Correlations and fluctuations

The degree to which the spatial arrangement of molecules in the jammed state deviates from randomness may be characterized by the *pair correlation function C_j*:

$$C_j \equiv \langle n_0\, n_j \rangle - \langle n_0 \rangle \langle n_j \rangle.$$ (7.27a)

Here n_j is the indicator function for the occupancy at site j, with $n_j = 0$ if site j is vacant and $n_j = 1$ if j is occupied. The correlation function can equivalently be written as[3]

$$C_j = \langle e_0\, e_j \rangle - \langle e_0 \rangle \langle e_j \rangle = P_j - E_1^2,$$ (7.27b)

where $e_j = 1 - n_j$ is the indicator function for empty sites, and $P_j \equiv \langle e_0\, e_j \rangle$. It is preferable to work with e_j and P_j because they are directly related to the empty interval probabilities.

---

[3] The constant and linear terms cancel in $C_j = \langle (1 - e_0)(1 - e_j) \rangle - \langle 1 - e_0 \rangle \langle 1 - e_j \rangle$.

The quantity P_j may be represented as the probability for the configuration

$$P_j = \mathcal{P}[\circ \underbrace{\times \cdots \times}_{j-1} \circ], \tag{7.28}$$

where $\times$ denotes a site with an unspecified state.

To determine the correlation function, we need the probabilities for the more general class of configurations that consist of empty intervals of lengths $\geq i$ and $\geq k$ that surround $j-1$ unspecified sites:

$$E_{i,j,k} \equiv \mathcal{P}\Big[\underbrace{\circ \cdots \circ}_{i} \underbrace{\times \cdots \times}_{j-1} \underbrace{\circ \cdots \circ}_{k}\Big].$$

Then P_j in (7.28) simply equals $P_j = E_{1,j,1}$.[4] The probabilities $E_{i,j,k}$ satisfy a hierarchy of master equations that are structurally similar to (7.3):

$$\frac{dE_{i,j,k}}{dt} = -(i-1)E_{i,j,k} - E_{i+1,j,k} - E_{i+1,j-1,k}$$
$$- (k-1)E_{i,j,k} - E_{i,j,k+1} - E_{i,j-1,k+1}, \tag{7.29}$$

for $i, k \geq 1$ and $j \geq 2$. The terms in the first line account, respectively, for the adsorption of a dimer within the empty i-string that overlaps the left end of this string, and that overlaps the right end of this string. The terms in the second line account for the corresponding events for the empty k-string.

Equations (7.29) are soluble, from which we can ultimately deduce the correlation function. We start by generalizing the ansatz of Eq. (7.4) and write

$$E_{i,j,k}(t) = e^{-(i+k-2)t} P_j(t) \quad \text{for} \quad i, k \geq 1 \tag{7.30}$$

with the initial condition $E_{i,j,k}(t = 0) = P_j(t = 0) = 1$. This ansatz reduces Eqs (7.29) to

$$\frac{dP_j}{dt} = -2e^{-t}\left[P_j + P_{j-1}\right] \tag{7.31}$$

for $j \geq 2$; for $j = 1$, we have $P_1 = E_2$. To solve this equation, we introduce the integrating factor $I = \exp\left(\int 2e^{-t}\,dt\right) = \exp\left[2(1 - e^{-t})\right]$, which happens to coincide with E_1^{-1} in (7.5), to simplify (7.31) to

$$\frac{dQ_j}{dt} = -2e^{-t} Q_{j-1}, \quad \text{with} \quad P_j = E_1 Q_j.$$

Finally, we absorb the explicit time dependence into a rescaled time variable $d\tau = -2e^{-t}\,dt$ to give

$$\frac{dQ_j}{d\tau} = Q_{j-1}, \quad \text{with} \quad \tau = \ln E_1. \tag{7.32}$$

---

[4] Notice that $E_{i,1,k} = E_{i+k}$, where the latter is the previously introduced empty interval probability, and also that $E_{i,2,k} = \mathcal{P}[\circ \cdots \circ \times \circ \cdots \circ] = E_{i+1+k}$ for dimer adsorption, since a *single* site that is sandwiched between clusters of empty sites must also be empty.
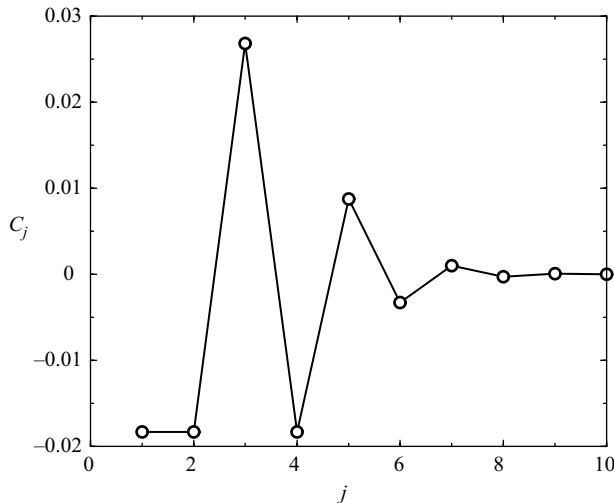
Fig. 7.7. The dependence of the correlation function C_j in Eq. (7.34) versus j.

This equation is subject to the initial condition $Q_j(0) = 1$ and the boundary condition $Q_1 = P_1/E_1 = E_2/E_1 = e^{-t}$, with the latter equality following from Eq. (7.5). Integrating Eqs (7.32) one by one starting with $Q_1 = 1 + \frac{1}{2}\tau$, gives $Q_2 = 1 + \tau + \frac{1}{2}\tau^2/2!$, $Q_3 = 1 + \tau + \tau^2/2! + \frac{1}{2}\tau^3/3!$, etc., and the general behavior is

$$Q_j = \sum_{k=0}^{j-1} \frac{\tau^k}{k!} + \frac{1}{2}\frac{\tau^j}{j!} \; .$$

Finally, we rewrite Q_j in terms of P_j and $\tau = \ln E_1$, and obtain the pair correlation function (Fig. 7.7)

$$C_j = P_j - E_1^2 = -E_1 \left[\frac{(\ln E_1)^j}{2 \cdot j!} + \sum_{k \geq j+1} \frac{(\ln E_1)^k}{k!} \right], \qquad j \geq 1. \qquad (7.33)$$

To derive this expression we use the expansion $e^x = \sum_{k \geq 0} x^k/k!$. In the jammed state $E_1(\infty) = e^{-2}$, so that

$$C_j \to -e^{-2}\frac{(-2)^{j-1}}{j!} \quad \text{as} \quad j \to \infty, \qquad (7.34)$$

which decays faster than the usual exponential decay that typifies correlations in equilibrium systems with short-range interactions, such as the hard-sphere gas. Notice also the positive (anti)correlation between sites with odd (even) separations.

The variance $\sigma^2 = \langle \rho^2 \rangle - \langle \rho \rangle^2$ is the standard measure for fluctuations in the coverage. From the definition $\rho = L^{-1} \sum_i n_i$ with L the system size, we write the variance in terms of the pair correlation functions $\langle n_i n_j \rangle$:

$$\sigma^2 = L^{-2} \sum_{i,j} [\langle n_i n_j \rangle - \langle n_i \rangle \langle n_j \rangle]. \tag{7.35}$$

We now write the correlation function as $C_k \equiv \langle n_i n_{i+k} \rangle - \langle n_i \rangle \langle n_{i+k} \rangle$, using Eq. (7.27a), with the index k any positive or negative integer. Taking the thermodynamic limit, summing over one of the two indices, and using $\sum_i \equiv L$, we find

$$\sigma^2 = L^{-1} \sum_{k=-\infty}^{\infty} C_k. \tag{7.36}$$

Hence, density correlations and density fluctuations are directly related. For systems with short-range correlations, as in adsorption, the fluctuations vanish in the thermodynamic limit and the density becomes a deterministic quantity. As an example, for dimer adsorption, we substitute (7.33) in (7.36) to give the variance $\sigma^2 = 2E_1^2 \ln(1/E_1)/L$, which equals $4e^{-4}L^{-1}$ in the jammed state (problems 7.10 and 7.11).

7.3 Adsorption in higher dimensions

Most physical applications of irreversible adsorption occur on two-dimensional substrates, where a natural starting point is the adsorption of elementary extended objects such as disks, rectangles, and sticks as a preliminary for understanding the adsorption of real particles, such as proteins and colloids. To get a sense for numerical values, the jamming coverages for the random sequential adsorption of various elementary objects in two dimensions are listed in Table 7.1. An exact analysis of adsorption is generally not possible in higher dimensions. Nevertheless, the evolution of the coverage in arbitrary dimensions has the same qualitative features as in one dimension. Namely, relaxation to the jamming density is exponential in time on discrete substrates and algebraic in time on continuous substrates. In the remainder of this section, we discuss simple examples that illustrate these features.
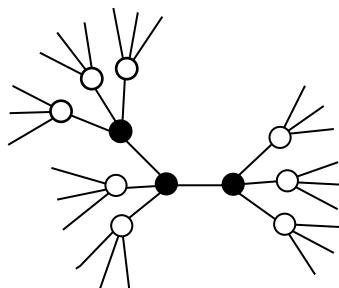
Discrete substrates

Adsorption is exactly soluble for one special high-dimensional substrate – the Cayley tree – in which each site is connected to exactly z other sites in a tree structure. Dimer adsorption on the Cayley tree means that the dimer can land equiprobably on any two neighboring unoccupied sites in the tree. Now the fundamental quantities are not empty intervals of length $\geq m$, but rather *empty connected clusters* of $\geq m$ sites that we again write as E_m.

Object	Substrate	ρ_{jam}
unoriented dimers	square lattice	0.9068
2×2 squares	square lattice	0.74788
aligned squares	continuum	0.562009
disks	continuum	0.5472

Table 7.1. Jamming coverages for various objects in two dimensions, obtained by numerical simulations.

Fig. 7.8. A cluster of $m = 3$ sites (bullets) and 8 perimeter sites (circles) on a Cayley tree with coordination number $z = 4$.

Because these clusters have a tree structure, it is straightforward to count the "boundary" configurations that enter into the evolution equations for E_m (Fig. 7.8). This feature allows us to solve adsorption on the Cayley tree.

The probability E_m that all sites in an empty m-cluster remain empty during dimer adsorption now satisfies (compare with Eq. (7.3))

$$\frac{dE_m}{dt} = -(m-1)E_m - \big[(z-2)m+2\big]E_{m+1}, \tag{7.37}$$

for $m \geq 1$, with the initial condition $E_m(0) = 1$. The first term on the right accounts for events in which the dimer adsorbs somewhere within the cluster. The number of available locations for such "interior" adsorption events is just the number of bonds in the cluster. Since every cluster has a tree topology, the number of bonds in the cluster is $m - 1$. The second term accounts for adsorption events in which one site of the dimer lands within an m-cluster and one site lands outside. The number of ways that such an event can occur equals the number of cluster perimeter sites – sites that adjoin the cluster, but are not part of the cluster itself (Fig. 7.8). For an m-cluster, the number of perimeter sites is $(z-2)m+2$, *independent* of its topology. The initial site has z perimeter sites and each successive site leads to a net gain in the perimeter by $z - 2$. Thus the total number of perimeter sites is $z + (m-1)(z-2) = (z-2)m+2$.

To solve Eqs (7.37), we apply the exponential ansatz $E_m(t) = \Phi(t)\left[\varphi(t)\right]^{m-1}$, with $\varphi(0) = \Phi(0) = 1$ to match the initial condition (see highlight on page 215). Then the hierarchy of equations (7.37) reduces to the pair of coupled equations

$$\frac{d\varphi}{dt} = -\varphi - (z-2)\varphi^2, \qquad \frac{d\Phi}{dt} = -z\,\varphi\,\Phi.$$

The (Bernoulli) equation for φ may be transformed to the soluble linear equation $du/dt = u + (z-2)$ by the transformation $u = 1/\varphi$, from which $\varphi = e^{-t}\left[(z-1) - (z-2)e^{-t}\right]^{-1}$. Using this result in the equation for Φ gives $\Phi(t) = \left[(z-1) - (z-2)e^{-t}\right]^{-z/(z-2)}$. Consequently, the empty cluster probabilities are

$$E_m(t) = e^{-(m-1)t}\left[(z-1) - (z-2)\,e^{-t}\right]^{-m-2/(z-2)}. \tag{7.38}$$

Finally, from the density $\rho(t) = 1 - E_1$, the approach to the jamming coverage is exponential in time, with jamming coverage

$$\rho_{\text{jam}} = 1 - (z-1)^{-z/(z-2)}. \tag{7.39}$$

In the limit $z \downarrow 2$, the one-dimensional result, $\rho_{\text{jam}} = 1 - e^{-2}$, is recovered (problem 7.14), while for large z, the uncovered fraction is inversely proportional to the coordination number, $1 - \rho_{\text{jam}} \sim z^{-1}$. The Cayley tree actually provides a reasonable approximation for ρ_{jam} for a regular lattice with the same coordination number as the tree. For example, when $z = 4$, dimer adsorption on the Cayley tree gives $\rho_{\text{jam}} = 8/9$, while for the square lattice, $\rho_{\text{jam}} \doteq 0.9068$.

A second look at exponential solutions (continued from page 139, Section 5.2)

Consider a master equation of the generic form

$$\frac{dE_m}{dt} = \lambda(m + \alpha)E_m + \mu(m + \beta)E_{m+1}, \tag{7.40}$$

that encompasses Eq. (7.37). Again, the exponential ansatz provides an easy route to the solution. Let's assume that E_m has the form

$$E_m = \Phi(t)[\varphi(t)]^m. \tag{7.41}$$

Substituting into Eq. (7.40) and then dividing by E_m gives

$$\frac{\dot{\Phi}}{\Phi} + m\frac{\dot{\varphi}}{\varphi} = \lambda(m + \alpha) + \mu(m + \beta)\varphi.$$

As in the exponential ansatz given in Section 5.2, the above equation divides naturally into terms linear in m and terms independent of m. From the linear terms, we have

$$\frac{\dot{\varphi}}{\varphi} = \lambda + \mu\varphi, \tag{7.42}$$

from which we obtain $\varphi(t)$. The m-independent terms yield

$$\frac{\dot{\Phi}}{\Phi} = \lambda\alpha + \mu\beta\varphi, \tag{7.43}$$

which then gives $\Phi(t)$.

The essential point is that the factor φ^m in the ansatz (7.41) separates the initial set of equations (7.40) into only two equations: one linear in m and one independent of m.

Apart from the Cayley tree solution, there are essentially no exact results about the evolution of the coverage for adsorption in greater than one spatial dimension. It is possible, however, to give a heuristic argument that the relaxation to the jamming coverage is exponential in time:

$$\rho_{\text{jam}} - \rho(t) \sim e^{-\lambda t}. \tag{7.44}$$

The behavior arises for the adsorption of arbitrarily shaped objects on discrete substrates in any dimension. We now discuss the simplest example of the adsorption of dimers on the square lattice.

Example 7.2. *Dimer adsorption on the square lattice*. Consider the state of the lattice at long times, when available spaces that can accommodate additional dimers are few and far between. These target regions are small contiguous clusters of unoccupied sites: dimers, trimers (both linear and bent), 2×2 squares, etc. To determine the rate at which these "lattice animals" get filled, we need the probabilities that these various configurations are empty. A simplifying feature at long times is that only the smallest empty lattice animals remain, and their asymptotic decay is dominated by the adsorption of dimers *inside* the animal. Thus for dimer adsorption on the square lattice, the probabilities of the lowest-order configurations (dimers, trimers, and 2×2 squares) evolve according to

$$\frac{d}{dt}\mathcal{P}[\circ\circ] \simeq -\mathcal{P}[\circ\circ], \qquad \frac{d}{dt}\mathcal{P}[\circ\circ\circ] \simeq -2\,\mathcal{P}[\circ\circ\circ], \qquad \frac{d}{dt}\mathcal{P}\left[\begin{smallmatrix}\circ\circ\\\circ\circ\end{smallmatrix}\right] \simeq -4\,\mathcal{P}\left[\begin{smallmatrix}\circ\circ\\\circ\circ\end{smallmatrix}\right].$$

Here, $\mathcal{P}[\,\cdot\,]$ denotes the likelihood of a configuration, and the numerical prefactor counts the number of ways that a dimer can adsorb within the cluster. The density of these configurations therefore decays with time as

$$\mathcal{P}[\circ\circ] \sim e^{-t}, \qquad \mathcal{P}[\circ\circ\circ] \sim e^{-2t}, \qquad \mathcal{P}\left[\begin{smallmatrix}\circ\circ\\\circ\circ\end{smallmatrix}\right] \sim e^{-4t}. \tag{7.45}$$

Generally, the probability that a given lattice animal is empty decays as $\mathcal{P}(t) \sim e^{-\lambda t}$, where λ counts the number of ways that a dimer can adsorb within a particular lattice animal. The coverage is determined by the lowest-order equation $d\rho/dt \sim -2\mathcal{P}[\circ\circ]$, so that $\rho_{jam} - \rho(t) \sim e^{-t}$.

Continuous substrates
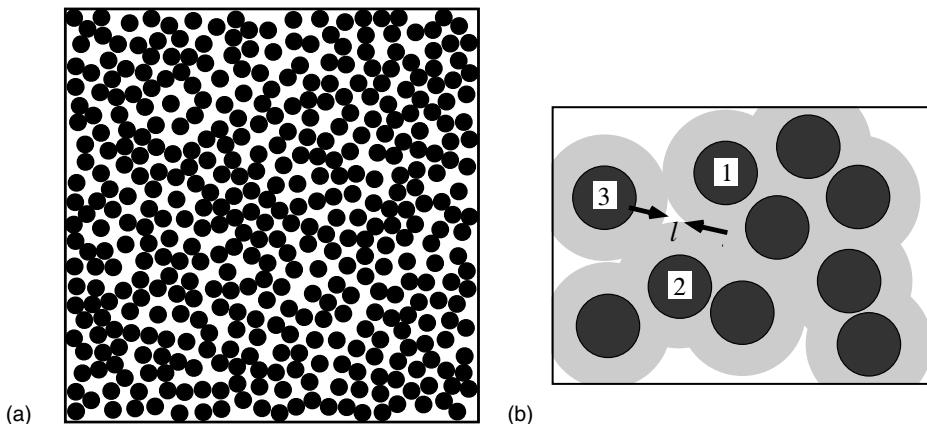
The kinetics of adsorption on continuous substrates is quite different in character than on discrete substrates. The main new feature is that gaps between adjacent adsorbed objects can be arbitrarily small (see Fig. 7.9), ultimately leading to an algebraic relaxation of the density,

$$\rho_{jam} - \rho(t) \sim t^{-\sigma} , \tag{7.46}$$

with σ the *jamming exponent*. This slow approach to the jamming density compared to discrete substrates also arises for car parking in one dimension (Eq. (7.16)), where the gaps between parked cars can also be arbitrarily small.

Let's now study the kinetics of irreversible adsorption of disks in two dimensions. Around each adsorbed disk there is an "exclusion zone" whose radius is twice that of the disk (Fig. 7.9(b)). If the center of an incident disk lies within the exclusion zone of any already adsorbed disk, the adsorption attempt fails. Therefore a disk can adsorb only if its center lies in the complement of the union of all excluded zones. We define these eligible portions of the substrate as "target zones." As the substrate approaches jamming, there will be only a few tiny and independent "target zones." In a jammed configuration, no target zones remain even though the adsorbed particles do not completely cover the substrate. The evolution of the coverage can be determined from the evolution of these target zones.

When, for example, disks 1 and 2 have already adsorbed and disk 3 is adsorbed next, a small target zone is created between these three disks (Fig. 7.9(b)). Let ℓ be the linear

(a)   (b)

Fig. 7.9.   (a) A jammed state for random sequential adsorption of unit-radius disks in a 50×50 area. (b) Illustration of a target zone (white), exclusion zones (shaded), and the adsorbed disks (dark) within a small cluster. In the long-time limit only arc-triangular target areas, such as the tiny one of linear size ℓ, exist.

dimension of this target zone. By construction, ℓ is uniformly distributed between some upper limit down to zero. Let $c(\ell, t)$ be the density of target zones of linear size ℓ. Because the area of such a target zone is quadratic in ℓ, the density of targets of linear size ℓ obeys $dc/dt \sim -\ell^2 c$, leading to

$$c(\ell, t) \sim e^{-\ell^2 t}. \tag{7.47}$$

The adsorption rate is just proportional to the area fraction of the target zones:

$$\frac{d\rho}{dt} \sim \int_0^\infty \ell^2 \, c(\ell, t) \, d\ell \sim \int_0^\infty \ell^2 \, e^{-\ell^2 t} \, d\ell \sim t^{-3/2}. \tag{7.48}$$
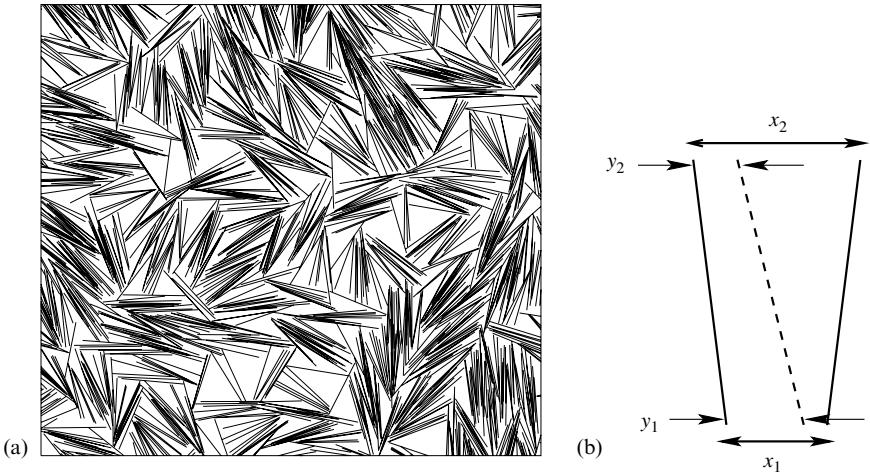
Therefore, the deviation of the substrate coverage from its jamming value is $\rho_{\text{jam}} - \rho(t) \sim t^{-1/2}$; that is, the jamming exponent $\sigma = \frac{1}{2}$. The salient point is that target zones can be arbitrarily small, and the superposition of exponential decays for each such target zone leads to a power-law decay (problem 7.15). This heuristic approach can be straightforwardly extended to arbitrary spatial dimension d. Now the area of a target zone of linear dimension ℓ scales as ℓ^d, and the corresponding density of target zones of linear dimension ℓ scales as $c(\ell, t) \sim e^{-\ell^d t}$. The analog of Eq. (7.48) then gives the jamming exponent $\sigma = 1/d$ in d dimensions. Thus relaxation slows down as the dimension increases.

Needles

An intriguing example of irreversible adsorption is the deposition of identical, randomly oriented, unit-length needles. The density of needles per unit area increases indefinitely with time and the chief issue is to determine the rate of this increase. Early deposition attempts are almost always successful because the substrate is nearly empty. Hence the number of adsorbed needles initially grows linearly with time and these initial needles have little orientational or positional order. However, when the needle density becomes of the order of 1, adsorbed needles strongly constrain the position and the orientation of subsequent adsorption events. In the long-time limit, domains form in which neighboring needles are nearly aligned *and* positionally ordered (Fig. 7.10). The orientation of each domain is fixed by the first few adsorption attempts, and the result is orientational order at small scales but not at large scales. Once domains are well defined, most adsorption attempts fail and the number of adsorbed needles grows sublinearly with time.

The formation of nearly aligned domains considerably simplifies the analysis of the long-time behavior. For a new needle to adsorb between two nearby needles, it must be closely aligned with them, both orientationally *and* positionally. At long times, the typical target area is approximately a unit-height trapezoid, with widths x_1 and $x_2 \approx x_1$ (Fig. 7.10(b)). A new needle may adsorb with its bottom end in the range $0 < y_1 < x_1$ and its top end in the range $0 < y_2 < x_2$. Such an adsorption event divides the trapezoid into two smaller trapezoids with base widths y_1, y_2 and $x_1 - y_1, x_2 - y_2$, thus defining a planar fragmentation process (see Section 6.2).

Using the equivalence to trapezoidal fragmentation and the corresponding master equation approach, the density $c(x_1, x_2)$ of trapezoids with widths (x_1, x_2) evolves

Fig. 7.10. (a) A long-time configuration of random sequential adsorption of needles in two dimensions. (b) The local geometry of a single needle adsorption event and its equivalence to the fragmentation of trapezoids.

according to

$$\frac{\partial c(x_1,x_2)}{\partial t} = -x_1 x_2\, c(x_1,x_2) + 2 \int_{x_1}^{\infty} \int_{x_2}^{\infty} c(y_1,y_2)\, dy_1\, dy_2. \tag{7.49}$$

The loss term is proportional to the total number of ways that each end of the needle can be placed along the bases of the trapezoid. The gain term accounts for the two ways in which the breakup of a larger trapezoid creates a trapezoid with base widths x_1, x_2. The Mellin transform $c(s_1,s_2) \equiv \iint x_1^{s_1-1} x_2^{s_2-1} c(x_1,x_2)\, dx_1\, dx_2$ evolves according to

$$\frac{\partial c(s_1,s_2)}{\partial t} = \left(\frac{2}{s_1 s_2} - 1\right) c(s_1+1, s_2+1). \tag{7.50}$$

As in rectangular fragmentation, there is again an infinite family of hidden conservation laws defined by $s_1^* s_2^* = 2$. If we assume that the Mellin transform has the algebraic time dependence

$$c(s_1,s_2) \sim t^{-\alpha(s_1,s_2)},$$

then Eq. (7.50) gives the recursion $\alpha(s_1+1, s_2+1) = \alpha(s_1,s_2) + 1$. Using this recursion and the condition $\alpha(s_1^*, s_2^*) = 0$ along the parabola $s_1^* s_2^* = 2$, the exponent $\alpha(s_1, s_2)$ may be determined by the same reasoning as that given in Section 6.2, and the result is

$$\alpha(s_1, s_2) = \frac{s_1 + s_2 - \sqrt{(s_1 - s_2)^2 + 8}}{2}. \tag{7.51}$$

One striking consequence of this formula is that the number density of needles varies *sublinearly* in time with *irrational* exponent

$$N \sim t^{\nu} \qquad \text{with} \quad \nu = \alpha(1, 1) = \sqrt{2} - 1. \tag{7.52}$$

As more needles are adsorbed, they become increasingly aligned. We quantify this alignment by $\langle \mathbf{n} \cdot \mathbf{n}' \rangle = \langle \cos \theta \rangle$, where $\mathbf{n}$ and $\mathbf{n}'$ are unit vectors along each needle and θ is the angle between them. This angle is related to the base length of the trapezoid defined by two needles via $\theta = |x_1 - x_2|$ for $x_1, x_2 \ll 1$. Thus the orientational correlation function $\langle \cos \theta \rangle$ is related to the moments of the trapezoid dimensions through

$$1 - \langle \cos \theta \rangle \sim \langle (x_1 - x_2)^2 \rangle = 2(\langle x^2 \rangle - \langle x_1 x_2 \rangle)$$

with $\langle x_1^2 \rangle = \langle x_2^2 \rangle \equiv \langle x^2 \rangle$. Using $\langle x^2 \rangle \sim t^{-\alpha(3,1)+\alpha(1,1)}$ we find that the orientational correlation function decays algebraically with another irrational exponent:

$$1 - \langle \cos \theta \rangle \sim t^{-\mu} \qquad \text{with} \qquad \mu = \alpha(3, 1) - \alpha(1, 1) = 1 + \sqrt{2} - \sqrt{3}. \tag{7.53}$$

Thus neighboring needles asymptotically become perfectly aligned.

The connection to trapezoidal fragmentation accounts, in an appealing way, for the subtle multiscaling properties and the irrational exponents that characterize needle adsorption. An important lesson is that a seemingly difficult problem – in this case needle adsorption – can be greatly simplified with the right geometric perspective. Once the dominant late-time dynamical mechanism is properly articulated, an exact asymptotic analysis of the problem becomes feasible.

7.4 Reversible adsorption

The physical situation in which an adsorbed molecule has a finite binding energy to a substrate motivates our study of *reversible* adsorption–desorption in which molecules adsorb with rate r_+ and desorb with rate r_-. As an everyday and macroscopic example, think of cars that are trying to park along a one-dimensional curb (Fig. 7.11). In keeping with common experience, if the rate at which cars leave (desorb) is small, the probability to find a parking spot also becomes small. Since the only relevant parameter is the ratio $r \equiv r_+/r_-$, we redefine the time as $t \to r_- t$ and write all formulae in terms of r.

Fig. 7.11. Adsorption–desorption of cars. Desorption attempts are always allowed (solid arrow), while the adsorption attempt shown (dashed arrow) fails because of insufficient parking space.

Example 7.3. *Reversible monomer adsorption.* As a preliminary, let's consider the situation in which monomers adsorb onto and desorb from single substrate sites. The problem is elementary because there is none of the frustration associated with car parking. The density of adsorbed particles ρ satisfies the Langmuir equation

$$\frac{d\rho}{dt} = -\rho + r(1 - \rho), \tag{7.54}$$

from which the relaxation to the steady state is exponentially fast:

$$\rho(t) = \rho_\infty + (\rho_0 - \rho_\infty) e^{-t/\tau} . \tag{7.55}$$

Here the relaxation time is $\tau = (1 + r)^{-1}$ and the final coverage is $\rho_\infty = r/(1 + r)$. Notice that when $r \to \infty$, the asymptotic coverage is $1 - \rho_\infty \simeq 1 - r^{-1}$, and this limit is reached in a time of the order of r^{-1}.

In reversible car parking, we seek the evolution of the coverage for unit-length cars that park with rate r in a continuous one-dimensional parking lot and depart at rate 1. Now, it is more convenient to use the density of voids of length x at time t, $V(x, t)$, rather than the density of empty intervals $E(x, t)$. Physically, the zeroth and first moments of the void density equal the fraction of occupied space and the fraction of empty space, respectively,

$$\rho = \int_0^\infty V(x) \, dx \qquad \text{and} \qquad 1 - \rho = \int_0^\infty x \, V(x) \, dx. \tag{7.56}$$

The first condition originates from the one-to-one correspondence between voids and parked cars, so that the density of voids of any size equals the car density.

For $x < 1$, the master equation for the void density is

$$\frac{\partial V(x, t)}{\partial t} = -2V(x, t) + 2r \int_{x+1}^\infty V(y, t) \, dy, \qquad x < 1. \tag{7.57a}$$

The first term on the right accounts for the loss of an x-void because of the desorption of a car at either end of the void. The second term accounts for the creation of a void of length x from a larger void by adsorption (problem 7.16). For voids of length $x > 1$, the evolution equation contains two additional terms:

$$\frac{\partial V(x, t)}{\partial t} = -2V(x, t) + 2r \int_{x+1}^\infty V(y, t) \, dy$$
$$- r(x - 1)V(x, t) + \frac{1}{\rho} \int_0^{x-1} V(y, t)V(x - y - 1, t) \, dy, \quad x > 1. \tag{7.57b}$$

The third term on the right accounts for the disappearance of a void due to adsorption events in the interior of the void. The disappearance rate is proportional to the amount of space available for adsorption, $x - 1$. The convolution term represents the creation of an x-void when a unit-length car leaves a parking spot that has an empty space of length y at one end of the car and a space $x - y - 1$ at the other end. The correct way to express the probability

for this composite event is through the two-void density. However, this description is not closed; the void evolution couples to a two-void density, and then the two-void function couples to three-void densities, *ad infinitum*. To break this hierarchy in the simplest manner, we assume that neighboring voids are uncorrelated and write the two-void density (voids of length y and $x - y - 1$ that surround a car) as the product of single-void densities. The factor ρ in the denominator gives the proper normalization. One can check that the master equations conserve the integral $\int (x + 1) V(x) \, dx = 1$, in accordance with (7.56).

Equilibrium steady states

While evolution equations such as (7.57a) and especially (7.57b) appear intimidating, they admit a simple exact solution. Notice that Eq. (7.57a) relates the void density to its spatial integral. This fact suggests the exponential solution $V(x) = A \, e^{-\alpha x}$. The two normalization conditions in (7.56) dictate that the prefactor is $A = \rho^2/(1-\rho)$ and the decay constant is $\alpha = \rho/(1 - \rho)$. Next, we observe that Eq. (7.57a) admits a steady-state solution when the decay constant satisfies $\alpha \, e^{\alpha} = r$. Finally, we substitute $V(x) = Ae^{-\alpha x}$, with A and α given above, in the convolution term and require that the two additional terms balance. This requirement again leads to $\alpha = \rho/(1 - \rho)$. Therefore the void density has the exponential form

$$V(x) = \frac{\rho^2}{1 - \rho} \, e^{-\rho x/(1-\rho)} \tag{7.58}$$

for all x. Remarkably, the form of the void density remains the same even though the underlying evolution equations (7.57a)–(7.57b) distinguish between voids that are larger than the car size and voids that are smaller. Moreover, gaps of all sizes are possible, in contrast to irreversible adsorption, where large gaps do not exist in the final state. Reversibility enables the system to explore the entire configurational phase space. In this sense, we may view the void density (7.58) as the equilibrium distribution. Finally, this equilibrium distribution also gives an *a posteriori* justification to our assumption that there are no correlations between adjacent voids (problems 7.16 and 7.17).

Thus the equilibrium density is found from the transcendental equation $\alpha \, e^{\alpha} = r$ with $\alpha = \rho/(1 - \rho)$. In the interesting limit of large adsorption rate,

$$\rho(r) \simeq 1 - (\ln r)^{-1} \, . \tag{7.59}$$

Thus r must be astronomically large to achieve a high density. For example, to reach $\rho = 0.99$, we must have $r = e^{100}$. Finally, notice that the behavior in the limit $r \to \infty$ is distinct from the situation when desorption is forbidden. Indeed, in the latter case, there is no mechanism to utilize too-small parking spaces and the jamming density is $\rho_{\text{jam}} \doteq 0.747597$.

Slow relaxation

The time dependence of the density is governed by

$$\frac{d\rho}{dt} = -\rho + r \int_1^\infty (x - 1) \, V(x) \, dx. \tag{7.60}$$

This generalized Langmuir equation (compare with Eq. (7.54)) has a simple interpretation: The first term on the right-hand side accounts for parked cars that desorb with rate ρ, thereby decreasing the density. The second term accounts for the increase in density by the parking of a car in a space of length $x > 1$. Note that (7.60) can also be obtained by integrating Eqs (7.57a)–(7.57b) over all lengths.

The interesting situation is $r \to \infty$, corresponding to the *desorption-controlled* limit, in which any vacated parking space is immediately refilled. In this limit, we can replace the exact equation (7.60) by an approximate equation that we justify as follows: On a time scale that is long compared to the desorption time, cars merely diffuse within their parking spaces as a result of desorption and immediate refilling, even though the identity of a car in a given parking spot changes. Thus the effect of the loss term in (7.60) is negated. Moreover, because of the rapid diffusion of cars within their parking spaces, the void density has the exponential form of Eq. (7.58). We therefore drop the loss term in Eq. (7.61) and also substitute the equilibrium void distribution (7.58) into the gain term to give

$$\frac{d\rho}{dt} = r(1 - \rho)\, e^{-\rho/(1-\rho)} \; . \tag{7.61}$$

Comparing Eqs (7.54) and (7.61), the factor

$$S(\rho) = e^{-\rho/(1-\rho)} \tag{7.62}$$

can be viewed as the probability that an adsorption event is successful – which becomes extremely unlikely as $\rho \to 1$. The density dependence of this effective sticking probability has a simple explanation. For a nearly full system, we write the density as $\rho = (1 + \langle x \rangle)^{-1}$ where $\langle x \rangle \ll 1$ represents the small bumper-to-bumper distance between neighboring parked cars. On a time scale of the order of one, a car leaves a parking space that in turn is immediately refilled. However, on a much larger time scale, desorption can lead to two parking spots opening up by a collective rearrangement of $N = \langle x \rangle^{-1} = \rho/(1 - \rho)$ parked cars. Such a fortunate event occurs if desorption and immediate readsorption (on a time scale of one) leads to the first car moving forward by $\langle x \rangle$, the second by $2\langle x \rangle$, the third by $3\langle x \rangle$, etc., until a new space of length one is eventually generated (Fig. 7.12). The probability for this cooperative rearrangement decays exponentially in the number of cars N, so that the effective sticking probability is $S \sim e^{-N}$, in agreement with (7.62). As the system approaches full coverage, the cooperative rearrangement involves a diverging number of cars so that the sticking probability becomes vanishingly small. To solve for
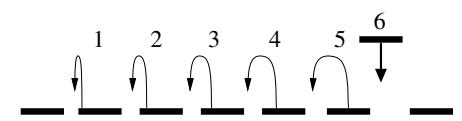
Fig. 7.12. Illustration of the six steps of a cooperative event in which five successive cars move slightly by desorption and subsequent immediate adsorption. Ultimately an additional parking space is created that is also immediately filled.

the time dependence of the car density in Eq. (7.61), we first write $g = (1 - \rho)^{-1}$. This replacement gives, to lowest order, $dg/d(rt) \sim e^{-g}$, whose solution is

$$1 - \rho(t) \sim \frac{1}{\ln(rt)}. \tag{7.63}$$

Thus adsorption of cars becomes extremely slow as the density increases because the sticking probability decays exponentially with the density.

The reversible car parking model provides helpful insight for the kinetics of granular compaction. Suppose that identical spherical glass beads are placed randomly into a jar one at a time. The density of filled space in this bead packing – approximately 0.58 – is known as the *random loose-packing density*. If this bead-filled jar is then gently tapped at a suitable intensity and frequency, the density will slowly increase and eventually reach the *random close-packing density* of approximately 0.64. The vibrations lead to occasional collective rearrangements that allows a bead to "park" in an empty interstitial space that was previously too small to accommodate the bead. As the density increases, rearrangements involve progressively larger spatial regions and become increasingly rare. The time scale over which this compaction occurs (Fig. 7.13) can be as long as months! However, this random close-packing density is still much smaller than the maximal density of $\pi/\sqrt{18} \doteq 0.7405$ for a periodic packing of spheres on a face-centered cubic lattice.

Diffusion-controlled adsorption

In reversible car parking, the separation of time scales between adsorption and desorption led to an effective model in which cars primarily diffused in their parking spaces. Whenever sufficient open space is created by collective rearrangements of parked cars, an additional car adsorbs immediately. Here, we study the situation where adsorbed molecules explicitly diffuse on the substrate and additional adsorption occurs whenever sufficient open space is created. Without loss of generality, we set the diffusion rate to one and the adsorption
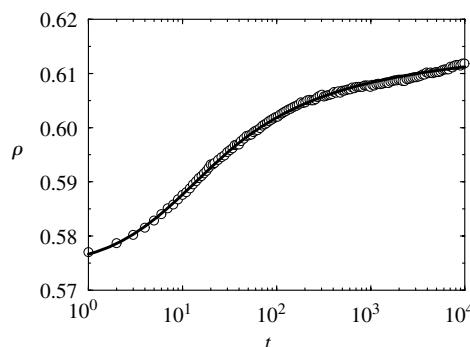
Fig. 7.13. Slow density relaxation in granular compaction (data from Nowak *et al., Phys. Rev.* E 57, 1971 (1998)). Shown is the density at a fixed height in a tapped granular column versus time. The solid line is the fit to an inverse logarithmic relaxation in (7.63).
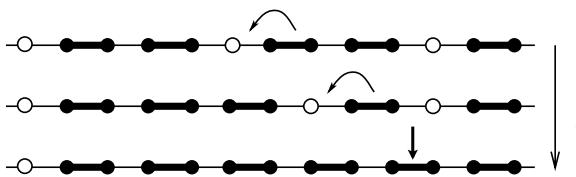
Fig. 7.14.   Mobile adsorbed dimers in one dimension. Each successive row shows the system after a single event. When two holes become adjacent, they are immediately occupied by a dimer.

rate to infinity. Thus whenever the configuration of adsorbed molecules permits another adsorption event, it occurs instantaneously, as illustrated in Fig. 7.14.

When dimers adsorb onto a one-dimensional lattice at an infinite rate, a jammed state is immediately reached in which all empty sites are isolated; we call such empty sites "holes." This jammed state is not stable because a dimer may hop to a neighboring site if there is a hole next to the dimer. As shown in Fig. 7.14, the hopping of a dimer in one direction corresponds to the effective hopping of the hole by two steps in the opposite direction. As a result of this hopping, two holes can occasionally become nearest neighbors. Such hole pairs are instantaneously and irreversibly filled by a dimer. This combination of diffusion and filling of adjacent hole pairs is equivalent to diffusion-controlled binary annihilation,

$$H + H \rightarrow \emptyset, \tag{7.64}$$
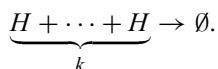
whose dynamics is well known (see Chapter 13 for a detailed presentation). Here the term "diffusion-controlled" signifies that diffusion limits the overall reaction, since annihilation occurs instantaneously whenever two reactants are in proximity. From the results that will be presented in Chapter 13, the density of holes ϕ decreases as

$$\phi \sim \begin{cases} t^{-1/2}, & d = 1, \\ t^{-1} \ln t, & d = 2, \\ t^{-1}, & d > 2, \end{cases} \tag{7.65}$$

where d is the spatial dimension of the substrate.[5] The change in behavior at two dimensions indicates that $d = d_c = 2$ is the critical dimension of diffusion-controlled annihilation.

We can similarly analyze the deposition of mobile k-mers. Now a hole hops by k lattice sites whenever an adsorbed molecule hops by one site. Whenever k holes are adjacent, an additional k-mer can be adsorbed. This diffusion and subsequent adsorption is equivalent to diffusion-controlled k-body annihilation,

$$\underbrace{H + \cdots + H}_{k} \rightarrow \emptyset.$$

---

[5] Adsorption is physically meaningful for spatial dimension $d = 1$ and $d = 2$, where all substrate sites are at the interface, but still remains mathematically well defined for all d.

For this reaction, the rate equation is now $d\phi/dt \sim -\phi^k$, and it applies for $k \geq 4$. The solution is[6]

$$\phi \sim t^{-1/(k-1)}, \qquad k \geq 4, \tag{7.66}$$

for all $d \geq 1$. The decrease of the hole density becomes progressively slower for longer molecules. In the limit $k \to \infty$, which corresponds to the car parking problem, the power-law relaxation (7.66) turns into the logarithmic relaxation law (7.63). Indeed, slow surface diffusion has the same effect as desorption and subsequent readsorption with a small displacement.

7.5 Polymer translocation

A particularly appealing application of adsorption kinetics is to the assisted translocation of a polymer through a small pore in a membrane. By thermal fluctuations alone, any biologically relevant polymer would pass through a pore too slowly to be relevant for cellular processes. One way to speed up translocation is to rectify the fluctuations so that backward motion is inhibited. As we now discuss, chaperone-assisted translocation provides just such a rectification mechanism. Suppose that the medium on the right side of the membrane contains a fixed density of "chaperones" that adsorb irreversibly onto unoccupied adsorption sites of the polymer with rate λ. If the chaperones are sufficiently large that they cannot pass through the pore, diffusion is rectified so that the polymer passes through the pore at a non-zero speed v.

We can understand the dependence of v on the control parameters D and λ by simple physical reasoning. Consider the limit of small adsorption rate, and let $\tau = 1/(\lambda \ell)$ be the typical time between adsorption events within the leftmost chaperone-free segment of length ℓ in the polymer (Fig. 7.15(c)). During this time, the polymer diffuses a distance $\ell \sim \sqrt{D\tau}$. Combining these two relations, we have $\tau \sim (D\lambda^2)^{-1/3}$ and $\ell \sim (D/\lambda)^{1/3}$. On

(a)   (b)   (c)
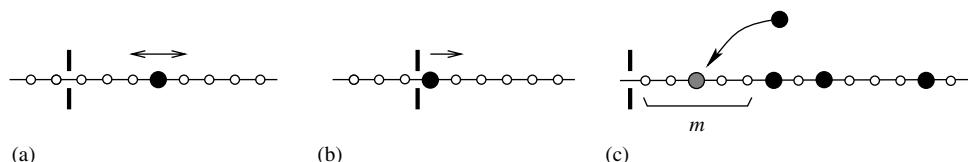
Fig. 7.15. Illustration of chaperone-assisted translocation. (a) The polymer can hop in either direction. (b) The polymer can hop only to the right because an adsorbed chaperone (large dot) is next to the pore (gap in the barrier) and too large to enter. (c) Adsorption of a new chaperone (shaded) within the leftmost chaperone-free segment.

[6] For trimers ($k = 3$), the critical dimension is $d_c = 1$ and there is a logarithmic correction, $\phi \sim t^{-1/2}\sqrt{\ln t}$ in one dimension. Otherwise, Eq. (7.66) holds for $d > 1$.

dimensional grounds, the translocation speed should therefore scale as

$$v \sim \frac{\ell}{\tau} \sim (D^2 \, \lambda)^{1/3} \, . \tag{7.67}$$

We now employ the empty interval method for irreversible adsorption that was developed in Section 7.1 to calculate the translocation speed v. For this purpose, we need to understand the dynamics of the leftmost chaperone-free segment $[1, m]$ of the polymer. There are two situations to consider. (i) The length m of this segment can change by ± 1 due to isotropic polymer hopping (with reflection at $m = 1$, see Fig. 7.15(b)); here we set the hopping rate to 1, corresponding to a "bare" polymer diffusion coefficient $D = 1$. (ii) Alternatively, m can change equiprobably to any value in $[1, m - 1]$ due to adsorption (Fig. 7.15(c)). The polymer therefore advances only when a chaperone is adjacent to the pore, while the polymer diffuses isotropically otherwise. Consequently, the speed equals the probability that the length of the leftmost chaperone-free interval equals 1. To compute this quantity, we need the probabilities E_m that an interval of length m to the right of the pore is chaperone-free:

$$E_m \equiv \mathcal{P}(\,|\, \underbrace{\circ \ldots \circ}_{m}\,) \, ;$$

here $|$ denotes the pore and $\circ$ an empty site. These are the analog of the empty interval probabilities defined in Eq. (7.2). The velocity is simply $v = 1 - E_1$; that is, when the site next to the pore contains a chaperone, the polymer translates with unit speed. We will also need the *segment probability* (see Eq. (7.7b))

$$S_m \equiv \mathcal{P}(\,|\, \underbrace{\circ \ldots \circ}_{m} \bullet\,) = E_m - E_{m+1} \, , \tag{7.68}$$

namely, the probability that the leftmost chaperone is at distance $m + 1$ from the pore; here $\bullet$ denotes the chaperone.

To compute E_1, the entire set of probabilities E_m is needed because finding E_1 requires E_2, which in turn is coupled to E_3, etc. The probabilities E_m evolve in time as

$$\frac{dE_m}{dt} = S_{m-1} - S_m - \lambda m E_m, \tag{7.69}$$

for $m \geq 1$. With the definition (7.68), the term S_{m-1} accounts for the creation of an empty interval of length m from an empty segment of length $m - 1$ as a result of the polymer hopping to the right. Similarly, an empty interval of length m is destroyed when the leftmost chaperone is a distance $m + 1$ from the pore and the polymer hops to the left. Hence the term $-S_m$. The term $-\lambda m E_m$ accounts for the loss of an m-interval because adsorption within this interval can occur in any one of m adsorption sites. By substituting (7.68) into (7.69), the empty interval probability evolves according to

$$\frac{dE_m}{dt} = E_{m-1} + E_{m+1} - 2E_m - \lambda m E_m, \tag{7.70}$$

for $m \geq 1$. Equation (7.70) is subject to the boundary condition $E_0 = 1$, which follows from the normalization condition $E_0 = \sum_{m \geq 0} S_m = 1$. Notice the segment probabilities

obey the same discrete diffusion that arises in the reaction $A + A \to 0$ with single-particle input (Eq. (13.35)).

In the long-time limit, a steady state is reached and the left-hand side of Eq. (7.70) vanishes,

$$E_{m-1} + E_{m+1} - (2 + m\lambda)E_m = 0. \tag{7.71}$$

We should solve this equation subject to the boundary condition $E_0 = 1$ and then extract the speed from $v = 1 - E_1$. Equation (7.71) is the discrete analog of the Airy equation, $E'' = \lambda x E$, and it admits a compact solution in terms of Bessel functions. Starting with the well-known identity for the Bessel functions,

$$J_{\nu-1}(z) + J_{\nu+1}(z) - \frac{2\nu}{z} J_\nu(z) = 0,$$

and comparing with Eq. (7.71), we must have

$$\frac{2\nu}{z} = 2 + m\lambda = 2\frac{m + 2/\lambda}{2/\lambda}.$$

Thus we identify $\nu = m + 2/\lambda$ and $z = 2/\lambda$. While there is only one condition to fix both ν and z, the choice is unique because the difference between two successive indices must be an integer. Finally, the requirement $E_0 = 1$ gives the solution

$$E_m = \frac{J_{m+2/\lambda}(2/\lambda)}{J_{2/\lambda}(2/\lambda)},$$

from which the translocation speed is

$$v = 1 - \frac{J_{1+2/\lambda}(2/\lambda)}{J_{2/\lambda}(2/\lambda)}. \tag{7.72}$$

The limiting behaviors of this translocation speed are instructive. In the small-λ limit, the index $\nu = m + 2/\lambda$ of the Bessel function becomes large and we use the asymptotic formula

$$J_\nu(\nu + x\nu^{1/3}) \simeq (2/\nu)^{1/3} \mathrm{Ai}(-2^{1/3}x), \tag{7.73}$$

that relates the Bessel function to another very useful special function, the Airy function. For large λ, (7.73) gives the power series representation

$$J_\nu(\nu + x) = (2/\nu)^{1/3} \mathrm{Ai}(0) - (2/\nu)^{2/3} \mathrm{Ai}'(0)\, x + \cdots, \qquad \nu \to \infty, \tag{7.74}$$

with $\mathrm{Ai}(0) = [3^{2/3}\Gamma(2/3)]^{-1}$ and $\mathrm{Ai}'(0) = -[3^{1/3}\Gamma(1/3)]^{-1}$. Substituting these results into (7.72), the leading behavior of the speed is (problem 7.18)

$$v \simeq \begin{cases} C\,\lambda^{1/3}, & \lambda \to 0, \\ 1 - \dfrac{1}{\lambda} + \dfrac{2}{\lambda^2} + \mathcal{O}(\lambda^{-3}), & \lambda \to \infty, \end{cases} \tag{7.75}$$

where the prefactor is $C = -\mathrm{Ai}'(0)/\mathrm{Ai}(0) = 3^{1/3}\Gamma(2/3)/\Gamma(1/3)$.

As a by-product, the probability $P_m = S_{m-1} = E_{m-1} - E_m$ that the leftmost chaperone is at m is given by

$$P_m = \frac{J_{m-1+2/\lambda}(2/\lambda) - J_{m+2/\lambda}(2/\lambda)}{J_{2/\lambda}(2/\lambda)} . \tag{7.76}$$

In the large-distance limit, it is simpler to study the continuum version of the steady-state equation (7.71), $E_{yy} = yE$, with $y = \lambda^{1/3}m$. We then employ the WKB approximation (see the highlight below) and obtain $E(y) \sim y^{-1/4} \exp\left(-\frac{2}{3}y^{3/2}\right)$. Finally, using $P = -\partial E/\partial m$ we find

$$P_m \sim \lambda^{-1/12}\, m^{1/4} \exp\left[-(2/3)\,\lambda^{1/2}\, m^{3/2}\right]. \tag{7.77}$$

Thus it is unlikely that the leftmost chaperone will be found at distances that are much larger than $\lambda^{-1/3}$, the average pore–chaperone distance.

The WKB method

The Wentzel–Kramers–Brillouin (WKB) method is a powerful technique to obtain the asymptotic solution of a differential equation near an irregular singular point. A prominent such example is the continuum limit of Eq. (7.71), namely, $E''(x) = \lambda xE$, or generically, $y'' = xy$. Such a form arises in the time-independent Schrödinger equation near a classical turning point. In the WKB approach we write y as the expansion $y = \exp[\phi_1(x) + \phi_2(x) + \cdots]$ and then solve for the functions ϕ_n recursively. To leading order, we obtain $(\phi_1')^2 + \phi_1'' \simeq x$. For large x, one term in this equation is negligible. Looking at both possibilities, one finds that the consistent choice is $(\phi_1')^2 \simeq x$, with $\phi_1'' \ll (\phi_1')^2$. There are two solutions to this equation, and the correct one is $\phi_1 = -\frac{2}{3}x^{3/2}$ which decays as $x \to \infty$. The next level of approximation gives $\phi_2' = -\frac{1}{4}x^{-1}$, from which the leading behavior is

$$y \sim x^{-1/4}\, e^{-2x^{3/2}/3} . \tag{7.78}$$

Continuing this method to the next level of approximation, all higher-order terms represent a vanishing correction to this leading behavior as $x \to \infty$.

7.6 Notes

Equation (7.1) that gives the jamming coverage in irreversible dimer adsorption was the first exact result in the field; it was derived by Flory [121] in the context of the cyclization reaction on a polymer chain. The corresponding jamming coverage for the car parking model was found by Rényi [122]. Subsequently, it was recognized that the kinetic approach provided a much simpler route to understand the basic properties of irreversible adsorption. Earlier exact results are reviewed in [123] and more recent reviews include [124–126]; these

contain many useful topics that were not covered in this chapter. The super-exponential decay of the correlation function in dimer adsorption is due to Monthus and Hilhorst [127].

While exact solutions of irreversible adsorption in one dimension were found in the early papers on the subject, not a single effective idea has been proposed on how to derive exact solutions in higher dimensions. What is needed is: first a method to exactly enumerate all possible configurations of connected clusters of any size on two-dimensional lattices (so-called "lattice animals"), and then to solve the coupled equations for the evolution of the probabilities that each lattice animal configuration is empty. However, as physicists, we feel content if we understand the asymptotic behavior of irreversible adsorption in higher dimensions. Such asymptotic arguments were presented in [128–130] for the adsorption of disks and aligned hypercubes. The discovery of the anomalous scaling of the coverage for the deposition of needles is due to Tarjus and Viot [131].

The reversible parking model [132] became popular because of its success in explaining the ultra-slow compaction of granular materials, with time scales of the order of months. The experimental data in Fig. 7.13 on slow compaction was presented in [133]. The connection between diffusion-controlled adsorption and diffusion-controlled annihilation is discussed in [134]. Chaperone-assisted polymer translocation is discussed in [135]; a more recent article, whose approach is close to our presentation, is [136]. The tome by Watson [137] contains (almost) everything known about Bessel functions. The WKB method is described in every book on quantum mechanics and many books on applied mathematics; an excellent presentation is given in Bender and Orszag [138].

7.7 Problems

7.1 For irreversible dimer adsorption, compute the total density of voids and the density of islands in terms of the empty interval probabilities.

7.2 Obtain the jamming coverage in restricted monomer adsorption in which monomers may adsorb only on an occupied site that is surrounded by two unoccupied sites. (*Hint*: Map this problem to a suitably defined dimer adsorption process.)

7.3 Suppose that dimers adsorb irreversibly onto a one-dimensional lattice in which each site is independently occupied with probability ρ_0 in the initial state. Compute the jamming coverage.

7.4 Using (7.11), determine the long-time behavior of $\rho_{\text{jam}} - \rho(t)$ for irreversible k-mer adsorption. Determine the limiting behavior of the coverage for $k \to \infty$.

7.5 Write and solve the equations for the evolution of irreversible dimer adsorption in terms of the void probabilities. Compare your equation with Eq. (6.2) that describes the random scission of an interval.

7.6 Derive Eq. (7.15) from (7.11) by taking the continuum ($k \to \infty$) limit of the former equation.

7.7 Solve the irreversible car parking model if initially there is a density λ of randomly distributed sizeless defects.

7.8 Generalize the recursive equation (7.17) for the total number J_L of jammed configurations to k-mer adsorption and determine the equation for z in $J_L \sim z^L$. Evaluate z for $k = 3, 4, 5$ and the leading asymptotic behavior of z for $k \to \infty$.

7.9 Use the finite-size correction given in Eq. (7.26) to extrapolate the thermodynamic jamming coverage in random sequential adsorption from finite-system simulations.

   (a) Use the efficient algorithm described in Section 7.1 to simulate irreversible dimer adsorption on a one-dimensional lattice of length L. Evaluate numerically the jamming coverage ρ_{jam} to the highest accuracy feasible by optimally tuning the number of realizations and the size of the system. (*Hint*: Perform separate extrapolations for L even and L odd.)

   (b) Obtain the jamming coverage for two-dimensional irreversible adsorption of dimers, where horizontal and vertical orientations are attempted with equal probabilities.

7.10 Compute the variance (7.36) by summing the correlation function (7.33).

7.11 Compute the structure factor $S(q) \equiv \sum e^{iqm} C_m$ from (7.33) for the jammed state of irreversible monomer adsorption.

7.12 Define the probabilities for filled strings of length m, F_m, and for islands of length m, I_m:

$$F_m \equiv \mathcal{P}[\underbrace{\bullet \cdots \bullet}_{m}], \qquad I_m \equiv \mathcal{P}[\circ \underbrace{\bullet \cdots \bullet}_{m} \circ].$$

Show that the island probability is the discrete second derivative of the filled string probability,

$$I_m = F_m - 2F_{m+1} + F_{m+2}.$$

7.13 For $m \leq 3$, use conservation statements to find the following relations between F_m and the empty interval probabilities E_m:

$$F_1 = 1 - E_1,$$
$$F_2 = 1 - 2E_1 + E_2,$$
$$F_3 = 1 - 3E_1 + 2E_2.$$

Notice that the general form of the last identity is

$$F_3 = 1 - 3\mathcal{P}[\circ] + 2\mathcal{P}[\circ\circ] + \mathcal{P}[\circ \times \circ] - \mathcal{P}[\circ \circ \circ],$$

but, for adsorption of dimers, $\mathcal{P}[\circ \times \circ] = \mathcal{P}[\circ \circ \circ]$ so that the three-body terms in this identity cancel.

For $3 < m < 7$, express F_m in terms of E_j and the probability for two disconnected empty intervals $E_{i,j,k}$. For F_5, for example,

$$F_5 = 1 - 5E_1 + 4E_2 - 2E_4 + E_5 + 2(1 - e^{-t})\Psi_3 + \Psi_4.$$

7.14 Verify that the expression (7.38) for the Cayley tree reduces to the expression (7.5) for dimer adsorption in one dimension by taking the limit $z \downarrow 2$ to recover the coverage for dimer adsorption.

7.15 Study the irreversible adsorption of oriented squares on a continuum planar substrate. Generalize (7.47) and (7.48) and determine the approach of the coverage to the jamming coverage.

7.16 Write the master equations, analogous to (7.57a)–(7.57b), for the void density in adsorption–desorption of k-mers in one dimension. Show that the void density is always exponential and find the equation governing the density. Obtain the density in the limit $r \to \infty$.

7.17 Calculate the equilibrium multi-void probability distribution $V_n(x_1, x_2, \ldots, x_N)$ that there are N voids with sizes $x_1, \ldots, x_N$ in reversible parking. Express this function in terms of the total void space and use the resulting expression to show the density fluctuations obey Gaussian statistics. Calculate the variance.

7.18 Consider the problem of chaperone-assisted translocation in which the chaperones are dimers that occupy two adjacent sites of the polymer. Compute the translocation speed for this case. More ambitiously, generalize to the case where the chaperone is a k-mer that occupies k consecutive sites on the polymer.

Spin dynamics

Two classical subjects in statistical physics are kinetic theory and phase transitions. The latter has been traditionally studied in the equilibrium framework, where the goal is to characterize the ordered phases that arise below the critical temperature in systems with short-range interactions. A more recent development has been to investigate how this order is formed *dynamically*. This and the following chapter focus on this dynamics.

8.1 Phenomenology of coarsening

The theory of phase transitions was originally driven by the goal of understanding ferromagnetism. The Ising model played a central role in this effort, and this model also provides a useful framework for investigating dynamics. The basic entity in the Ising model is a spin variable that can take two possible values, $s = \pm 1$, at each site of a lattice. A local ferromagnetic interaction between spins promotes their alignment, while thermal noise tends to randomize their orientations. The outcome of this competition is a disordered state for sufficiently high temperature, while below a critical temperature the tendency for alignment prevails and an ordered state arises in which the order parameter – the average magnetization – is non-zero.

Suppose that we start with an Ising model in an equilibrium disordered phase and lower the temperature. To understand how the system evolves toward the final state, we must endow this model with a dynamics and we also need to specify the quenching procedure. Quenching is usually implemented as follows:

- Start at a high initial temperature $T_i > T_c$, where spins are disordered; here T_c is the critical temperature.
- *Instantaneously* cool the system to a lower temperature T_f.

In various applications (ranging from the formation of glasses to the cooling of the universe) slow cooling is appropriate, but fast cooling is natural in many other settings. The lack of a temporal scale in instantaneous cooling is particularly useful, as it allows us to focus on the time scale associated with coarsening rather than the scale that is imposed by the quenching procedure.

The definition of the dynamics itself is quite flexible. Since atomistic spins are the basic objects, in principle we should use the Heisenberg or the Schrödinger quantum equations of motion. Such a fully microscopic theoretical treatment of a strongly interacting non-equilibrium quantum many-particle system is essentially hopeless. Already for a few dozen

such particles, a numerical treatment becomes extraordinarily challenging. Moreover, even for dilute classical gases, little can be understood in the framework of Newton's equations and a coarse-grained (and probabilistic) Boltzmann approach is necessary. Similarly for kinetic spin systems we need to devise less fundamental and more tractable frameworks.

A major simplification in studying the evolution of spin systems is that we can ignore quantum effects. The justification for this approximation is that phase transitions in ferromagnets occur far above the temperatures where quantum effects are important.[1] Within a classical description, the master equation is our primary analysis tool to treat many-body systems at a probabilistic level. Here, we apply this approach to models with simple microscopic evolution rules that drive systems toward equilibrium; we investigate these types of models from a more macroscopic (phenomenological) perspective in the next chapter.

For the evolution equations that will be outlined in these two chapters, little rigorous knowledge has been established. Despite the multitude of theories and paucity of firm results, what is known – from simulations, approximate and heuristic arguments, and exact analysis of tractable (usually one-dimensional) models – points to the validity of a simple *central dogma*. This dogma asserts that there are three broad classes of dynamical behavior for spin systems, depending on whether the final temperature is larger or smaller than the critical temperature.

1. *Supercritical dynamics, $T_f > T_c$.* The equilibrium disordered state corresponding to the final temperature T_f is reached exponentially quickly at a temperature-independent rate.
2. *Critical dynamics, $T_f = T_c$.* Critical dynamics applies only at one temperature value, T_c. There are as many types of critical dynamics as there are types of equilibrium critical behavior.
3. *Subcritical dynamics, $T_f < T_c$.* The dynamical behavior is universal and essentially the same as that for the zero-temperature dynamics.

Most studies have concentrated on zero-temperature dynamics because of its universal character, because it leads to crisp results, and because it is much simpler than finite-temperature dynamics. The most striking feature of zero-temperature dynamics (and subcritical dynamics generally) is the emergence of an intricate coarsening domain mosaic (Fig. 8.1). Such domain mosaics exhibit a remarkable degree of universality in their basic properties. These universal features can be summarized by the so-called *dynamical scaling hypothesis*.

1. *Single scale.* In the late stages of coarsening, the system is characterized by a single length scale $L(t)$; that is, if we rescale the spatial coordinates $\mathbf{x} \rightarrow \mathbf{x}/L(t)$, domain mosaics at different times will look (statistically) the same.
2. *Algebraic scaling.* The characteristic length scale $L(t)$ grows with time as t^z in the long-time limit; here z is the universal *dynamical exponent*.
3. *Universality of the dynamical exponent.* The exponent z is independent of the spatial dimensionality and of most system details, but is sensitive to conservation laws.

---

[1] Quantum fluctuation effects are relevant at zero temperature, where thermal fluctuations vanish. We do not discuss this burgeoning field of zero-temperature quantum phase transitions in this book. Even when we study zero-temperature dynamics, we do so within classical physics.
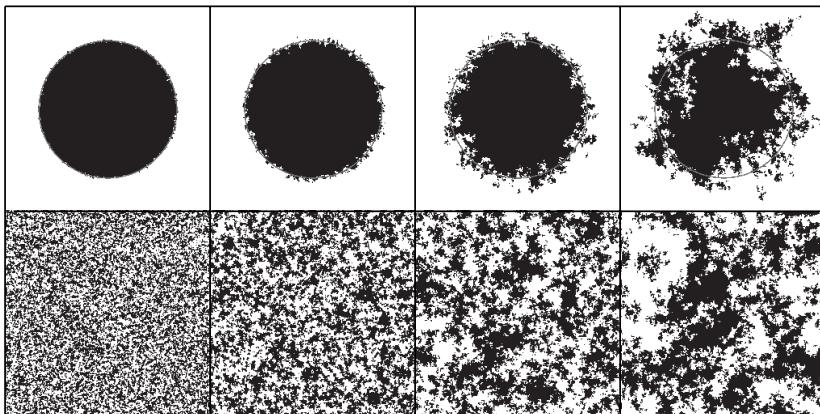
Evolution of the voter model on a 256 × 256 square lattice. The top row shows snapshots at times $t = 4, 16, 64$, and 256 starting with an initial circular bubble of one opinion of radius 180 in a sea of the opposite opinion. The bottom row shows the same evolution starting with a random initial condition with equal densities of the two opinion states. The figure is from Dornic *et al.*, *Phys. Rev. Lett.* 87, 045701 (2001); courtesy of I. Dornic.

The central dogma and the dynamical scaling hypothesis provide a comprehensive framework for understanding coarsening processes. The existence of such a framework is striking, since most non-equilibrium processes are beyond the reach of mathematical analysis. Before embarking on analysis, we remark that one-dimensional systems with a scalar order parameter provide the best opportunity to probe dynamical behavior quantitatively. We shall see that this is indeed the case, but the one-dimensional setting is pathological because the phase transition occurs at $T_c = 0$.[2] Thus it is not clear *a priori* if solutions to one-dimensional problems correspond to critical or subcritical dynamics. We shall see, however, that predictions based on solutions to the master equations of one-dimensional models give the same qualitative results as zero-temperature dynamics in higher dimensions. There are some exceptions to the general picture, however. In particular, Langevin dynamics predict a slow logarithmic coarsening which contradicts the dynamical scaling hypothesis; however, the existence of a single length scale still holds. There are also other complications that will be discussed later in this chapter; e.g. a system quenched to zero temperature may get trapped forever in a metastable state. The fundamental message, however, is that the central dogma and the dynamical scaling hypothesis are "morally" correct, although they are not literally correct in some special cases.

8.2 The voter model

We begin our study of spin dynamics with the voter model – essentially a toy spin system that has an advantage of being solvable in arbitrary spatial dimensions. This model describes consensus formation in a population of spineless individuals. Each individual is situated on

---

[2] Another such example is a system with a vector order parameter in two dimensions, where again $T_c = 0$.
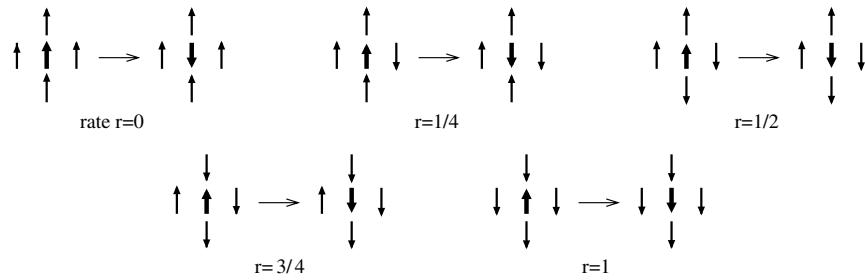
Fig. 8.2. **The flip rates in Eq. (8.1) for the central voter as a function of the state of its local neighborhood on the square lattice.**

the sites of a graph[3] and can be in one of q equivalent opinion states. An individual does not have any conviction and merely adopts the opinion of one of its neighbors in an update event. By this dynamics, a finite population of voters eventually comes to consensus in a time that depends on the system size, the spatial dimension, and the number of opinions q. Consensus in one of the q equivalent opinions is necessarily reached because this configuration is the *only* absorbing state of the stochastic dynamics. For simplicity, we deal with the two-state voter model henceforth. The evolution of the voter model is shown in Fig. 8.1 to illustrate how the system organizes into single-opinion domains as time increases.

At each lattice site $\mathbf{x}$, a voter is situated whose opinion, $s(\mathbf{x})$, can be in one of two states ± 1; we may view this opinion as an Ising-like variable. Each spin flips at a rate that equals the fraction of its neighbors in the opposite opinion state (Fig. 8.2). This flipping rule can be encoded by the rate $w_{\mathbf{x}}(\mathbf{s})$ at which the opinion $s(\mathbf{x})$ of the voter at $\mathbf{x}$ changes state:

$$w_{\mathbf{x}}(\mathbf{s}) = \frac{1}{2}\Big[1 - \frac{s(\mathbf{x})}{z} \sum_{\mathbf{y} \in \langle \mathbf{x} \rangle} s(\mathbf{y})\Big], \tag{8.1}$$

where the sum is over the nearest neighbors $\langle \mathbf{x} \rangle$ of site $\mathbf{x}$ and z is the lattice coordination number. The index $\mathbf{x}$ in the rate $w_{\mathbf{x}}(\mathbf{s})$ indicates that only the voter at site $\mathbf{x}$ changes opinion in an update event. Notice that the transition rate of a voter at $\mathbf{x}$ equals the fraction of disagreeing neighbors. As we shall see, this *proportionality* of the update rule is the primary reason why the voter model is soluble.

The full description of the voter model is provided by $P(\mathbf{s}, t)$, the probability that the set of all voters are in configuration $\mathbf{s}$ at time t. This probability distribution evolves according to[4]

$$\frac{dP(\mathbf{s})}{dt} = -\sum_{\mathbf{x}} w_{\mathbf{x}}(\mathbf{s})P(\mathbf{s}) + \sum_{\mathbf{x}} w_{\mathbf{x}}(\mathbf{s}^{\mathbf{x}})P(\mathbf{s}^{\mathbf{x}}). \tag{8.2}$$

---

[3] A graph – either finite or infinite – is a collection of vertices, some of which are joined by edges. For spatially regular infinite lattices, and generally in physics, the vertices are called sites and the edges are called bonds. We adopt this terminology and also mostly consider regular lattices in this chapter.

[4] In Eq. (8.2) we drop the time argument to make formulae less cluttered. Also, the notation $P(\mathbf{s}, t)$ is imprecise because the probability distribution does *not* explicitly depend on time. Rather, the distribution is a function of $\mathbf{s}$, which in turn depends on time, so that $P(\mathbf{s}[t])$ is a *functional* of $\mathbf{s}$. We shall avoid this pedantry and often suppress the time variable even for quantities that explicitly depend on time; e.g. in Eq. (8.5), we write $S(\mathbf{x})$ rather than $S(\mathbf{x}, t)$.

Here $\mathbf{s}^{\mathbf{x}}$ denotes the state that is the same as $\mathbf{s}$ except that the voter at $\mathbf{x}$ has changed opinion, that is, $\mathbf{s}^{\mathbf{x}}(\mathbf{y}) = s(\mathbf{y})$ when $\mathbf{y} \neq \mathbf{x}$ and $\mathbf{s}^{\mathbf{x}}(\mathbf{x}) = -s(\mathbf{x})$. In Eq. (8.2), the loss term accounts for all possible transitions out of state $\mathbf{s}$, while the gain term accounts for transitions to state $\mathbf{s}$ from states in which one spin differs from the configuration $\mathbf{s}$. Instead of solving for the full probability distribution, we focus on correlation functions, from which we can deduce the most important features of the voter model.

Simulating the voter model

An algorithmic definition of the voter model update rule is as follows:

1. Pick a random voter equiprobably, say at site $\mathbf{x}$.
2. The selected voter adopts the state of a randomly selected neighbor, say at site $\mathbf{y}$. (A voter thus changes opinion only when its interaction partner has the opposite opinion.)
3. Increment the time appropriately, $t \rightarrow t + \Delta t$.

These steps are repeated until either a specified time is reached or consensus is achieved. A natural choice for the time increment is $\Delta t = 1/N$ so that, when each voter has attempted to change its state once, on average, the time increases by one. An alternative that is closer to Eq. (8.2) is to choose a time increment that is drawn from an exponential random variable whose mean value is $1/N$.

  In a real simulation, one should not implement the above naive update recipe because it is hopelessly inefficient. As the population approaches consensus, it becomes progressively more likely to pick two voters in the same states, and the outcome of such a selection is to do nothing. Instead of simulating null events, one should implement an event-driven algorithm in which only "successful" oppositely oriented voters are picked. The average time increment for a successful event is $(2N_{+-})^{-1}$, where N_{+-} is the number of oppositely oriented connected voters.

Correlation functions

In principle, we can use the master equation (8.2) to derive closed equations for all moments of the probability distribution – namely, all many-body correlation functions of the form $S_{\mathbf{x},\dots,\mathbf{y}} \equiv \langle s(\mathbf{x}) \dots s(\mathbf{y}) \rangle$ where the angle brackets denote the average $\langle f(\mathbf{s}) \rangle \equiv \sum_{\mathbf{s}} f(\mathbf{s}) P(\mathbf{s})$. Let's begin by considering the simplest such correlation function, namely, the average opinion, or equivalently the one-point function, $S(\mathbf{x}) \equiv \langle s(\mathbf{x}) \rangle$. It is possible to obtain the evolution of this function, and indeed any many-body correlation function, directly from (8.2). This approach involves tedious bookkeeping that is prone to error, however. A simpler method is to note that, in an infinitesimal time interval Δt, the state of a given voter changes as follows:[5]

$$s(\mathbf{x}, t + \Delta t) = \begin{cases} s(\mathbf{x}, t), & \text{with probability } 1 - w_{\mathbf{x}}\Delta t, \\ -s(\mathbf{x}, t), & \text{with probability } w_{\mathbf{x}}\Delta t. \end{cases} \tag{8.3}$$

---

[5] We often write $w_{\mathbf{x}}$ instead of $w_{\mathbf{x}}(\mathbf{s})$; as mentioned in the previous footnote, the precise notation would be $w_{\mathbf{x}}(\mathbf{s}[t])$.

Since the opinion at $\mathbf{x}$ changes by $-2s(\mathbf{x})$ with rate $w_{\mathbf{x}}$, the average opinion evolves according to

$$\frac{dS(\mathbf{x})}{dt} = -2\langle s(\mathbf{x})w_{\mathbf{x}}\rangle. \tag{8.4}$$

Substituting the transition rate from (8.1) and using $[s(\mathbf{x})]^2 = 1$, the mean opinion in the voter model evolves as

$$\frac{dS(\mathbf{x})}{dt} = -S(\mathbf{x}) + \frac{1}{z}\sum_{\mathbf{y}\in\langle\mathbf{x}\rangle} S(\mathbf{y}), \tag{8.5}$$

where z is the coordination number of the lattice.

Mathematically, Eq. (8.5) is the same as the master equation for a continuous-time lattice random walk. As a result, the mean magnetization, $m \equiv \sum_{\mathbf{x}} S(\mathbf{x})/N$, plays the role of the total probability and hence is conserved. This same result follows by summing Eq. (8.5) over all sites. Notice that while the magnetization of a specific system *does* change in a single update event, the average over all sites and over all trajectories of the dynamics *is* conserved.

The consequences of this conserved magnetization are profound (Fig. 8.3). Consider a finite system with an initial fraction ρ of voters in the $+1$ state and a fraction $1-\rho$ in the -1 state, so that the initial magnetization is $m_0 = 2\rho - 1$. Ultimately, this voter system reaches consensus in which the state with magnetization $m = +1$ occurs with probability $E(\rho)$ and the state with $m = -1$ occurs with probability $1 - E(\rho)$. The final magnetization is

$$m_\infty = E(\rho) \times 1 + [1 - E(\rho)] \times (-1) = 2E(\rho) - 1,$$

which equals the initial magnetization $m_0 = 2\rho - 1$. To summarize, the probability of reaching $+1$ consensus is $E(\rho) = \rho$ and the probability of reaching -1 consensus is $1 - \rho$. These probabilities depend only on ρ; there is no dependence on the system size and the structure of the underlying graph.[6]
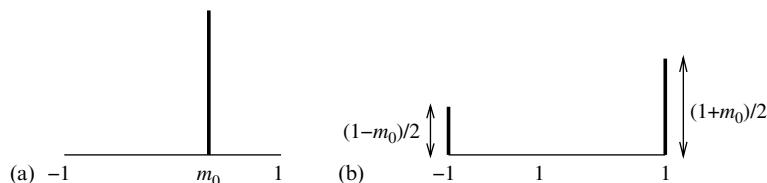
Illustration of magnetization conservation. (a) The initial state has magnetization m_0. (b) The final state is either that of positive consensus ($m = +1$) or negative consensus ($m = -1$) and magnetization conservation fixes the probabilities of achieving these two states.

---

[6] These statements do not apply if the underlying graph consists of disconnected pieces or if each site does not have the same number of neighbors.

Let's solve Eq. (8.5) subject to the initial condition $S(x, t = 0) = \delta_{x,0}$; that is, a single voter is in the $+1$ state in a background population of undecided voters. In one dimension,

$$\frac{dS(x)}{dt} = -S(x) + \frac{1}{2}\left[S(x-1) + S(x+1)\right]. \tag{8.6}$$

Using the results from the highlight on the Bessel function solution to this type of difference equation (page 14 in Section 2.1), we immediately obtain

$$S(x, t) = e^{-t} I_x(t). \tag{8.7}$$

The same approach works on higher-dimensional lattices. For the hypercubic lattice, for instance, we perform the multidimensional Fourier transform $P(\mathbf{k}, t) = \sum_{\mathbf{x}} P(\mathbf{x}, t)\, e^{i\mathbf{k}\cdot\mathbf{x}}$ and notice that the Fourier transform in each coordinate direction factorizes. Thus for the initial condition of a single $+1$ voter at the origin in a sea of undecided voters, the average opinion at site $\mathbf{x} = (x_1, \ldots, x_d)$ is

$$S(\mathbf{x}, t) = e^{-t} I_{\mathbf{x}}(t/d), \qquad I_{\mathbf{x}}(t/d) \equiv \prod_{i=1}^{d} I_{x_i}(t/d). \tag{8.8}$$

Thus the average opinion of the voter at the origin monotonically decays to zero, while the average opinion of any other voter initially rises and reaches a maximum value (that depends on its location $\mathbf{x}$) and then decays to zero. Asymptotically $S(\mathbf{x}, t) \simeq (d/2\pi t)^{d/2}$ for any $\mathbf{x}$. Thus a single voter relaxes to the average undecided opinion of the rest of the population.

While the above result is exact, it provides no information about how consensus is actually achieved. What we need is a quantity that quantifies the extent to which two distant voters agree. Such a measure is provided by the two-point correlation function, $S(\mathbf{x}, \mathbf{y}) = \langle s(\mathbf{x})s(\mathbf{y}) \rangle$. Proceeding in close analogy with Eq. (8.3) the two-point function evolves as

$$s(\mathbf{x}, t + \Delta t)s(\mathbf{y}, t + \Delta t) = \begin{cases} s(\mathbf{x}, t)s(\mathbf{y}, t), & \text{with probability } 1 - (w_{\mathbf{x}} + w_{\mathbf{y}})\Delta t, \\ -s(\mathbf{x}, t)s(\mathbf{y}, t), & \text{with probability } (w_{\mathbf{x}} + w_{\mathbf{y}})\Delta t. \end{cases} \tag{8.9}$$

That is, $s(\mathbf{x})s(\mathbf{y})$ changes by $-2s(\mathbf{x})s(\mathbf{y})$ if either of the voters at $\mathbf{x}$ or $\mathbf{y}$ changes state with respective rates $w_{\mathbf{x}}$ and $w_{\mathbf{y}}$, so that $S(\mathbf{x}, \mathbf{y})$ evolves according to

$$\frac{dS(\mathbf{x}, \mathbf{y})}{dt} = -2\langle s(\mathbf{x})s(\mathbf{y})(w_{\mathbf{x}} + w_{\mathbf{y}}) \rangle.$$

On a hypercubic lattice, the explicit form of the above equation is

$$\frac{dS(\mathbf{x}, \mathbf{y})}{dt} = -2S(\mathbf{x}, \mathbf{y}) + \frac{1}{2d} \sum_{\mathbf{x}' \in \langle \mathbf{x} \rangle} S(\mathbf{x}', \mathbf{y}) + \frac{1}{2d} \sum_{\mathbf{y}' \in \langle \mathbf{y} \rangle} S(\mathbf{x}, \mathbf{y}'). \tag{8.10}$$

In what follows, we discuss spatially homogeneous systems in which the correlation function depends only on the distance between voters: $S(\mathbf{x}, \mathbf{y}) \equiv G(\mathbf{x} - \mathbf{y})$. For such systems the last two terms on the right-hand side of Eq. (8.10) are identical and this equation reduces to (8.5) for the mean spin, apart from an overall factor of 2. However, we cannot merely use the solution for the mean spin because the two-point correlation function must also satisfy the boundary condition $G(\mathbf{0}, t) = 1$.

Continuum approach

The solution Eq. (8.10), subject to the boundary condition $G(\mathbf{0}, t) = 1$, is somewhat technical, so let's first try to extract the most interesting results in a simpler fashion, where "interesting" is usually a synonym for "asymptotic." When the time scale becomes large, the relevant spatial scale is also expected to be large.[7] Therefore we first explore the voter model in the continuum limit where (8.10) becomes the diffusion equation

$$\frac{\partial G}{\partial t} = D \nabla^2 G, \tag{8.11}$$

and the diffusion coefficient associated with this continuum limit is $D = 1/d$. For the undecided initial state in which each voter is independently ± 1 with equal probability, the initial condition is $G(r, t = 0) = 0$ for $r > 0$. On the other hand, each voter is perfectly correlated with itself, that is $S(\mathbf{x}, \mathbf{x}) = 1$. In the continuum limit, we must impose a lower cutoff $a > 0$ in the argument of the correlation function, so that the statement of perfect self-correlation becomes $G(a, t) = 1$. While the lattice system is certainly not isotropic, the correlation function $G(\mathbf{r})$ should become isotropic when the separation is much greater than the lattice spacing, $r \gg a$. Since the imposed initial condition $G(a, t) = 1$ is isotropic, the corresponding solution $G(r, t)$ of the continuum diffusion equation (which is rotationally invariant) is also isotropic. Thus we expect that the correlation function $S(\mathbf{x}, \mathbf{y})$ of the lattice system approaches $G(r, t)$ when $r = |\mathbf{x} - \mathbf{y}| \gg a$.

To understand how the correlation function evolves, it is expedient to work with $c \equiv 1 - G$ that also satisfies the diffusion equation, but with the initial condition $c(r > a, t = 0) = 1$ and the boundary condition $c(r = a, t) = 0$. That is, the absorbing point at the origin is replaced by a small absorbing sphere of non-zero radius a. One should think of a as playing the role of the lattice spacing – a non-zero radius is needed so that a diffusing particle can actually hit the sphere. We are therefore studying how an initially constant density profile evolves in the presence of a small absorbing sphere at the origin. This problem has already been analyzed by the quasi-static approximation (see Section 2.5). There we found that, in three dimensions, the concentration profile becomes stationary, $c(r) = 1 - a/r$; more generally, when $d > 2$, the concentration profile approaches the electrostatic solution, $c(r) = 1 - (a/r)^{d-2}$. In one and two dimensions, the concentration profile approaches the trivial solution $c = 1$. From Eq. (2.88) in Section 2.5, the asymptotic behavior of the

---

[7] We will see that for the voter model the relevant spatial scale indeed grows indefinitely when $d = 1, 2$. In higher dimensions ($d > 2$), this growth no longer occurs; nevertheless even for this situation the continuum approach provides valuable insights.

two-spin correlation function is given by:

$$G(r,t) \sim \begin{cases} 1 - \dfrac{r}{\sqrt{Dt}}, & d = 1 \quad \text{and} \quad 0 < r < \sqrt{Dt}, \\[2mm] 1 - \dfrac{\ln(r/a)}{\ln(\sqrt{Dt}/a)}, & d = 2 \quad \text{and} \quad a < r < \sqrt{Dt}, \\[2mm] \left(\dfrac{a}{r}\right)^{d-2}, & d > 2 \quad \text{and} \quad a < r. \end{cases} \tag{8.12}$$

Thus for $d \leq 2$ the correlation function at fixed separation r approaches 1, i.e. distant spins gradually become more correlated. This feature is a manifestation of *coarsening* in which the voters organize into single-opinion enclaves whose characteristic size increases with time. On the other hand, for $d > 2$ the voter model approaches a steady state and there is no coarsening in the spatial arrangement of the voters if the population is infinite.

There are two important consequences for the voter model that can be deduced from the correlation function. The first is the time dependence of the density of "interfaces," namely, the fraction ρ of neighboring voters that have opposite opinions. As we shall also use extensively later in this chapter, it is helpful to represent an interface as an effective quasi-particle that occupies the bond between two neighboring opposite-opinion voters. The density of domain walls characterizes the departure of the system from consensus. For nearest-neighbor sites $\mathbf{x}$ and $\mathbf{y}$, we relate the correlation function to the domain-wall density by

$$G(\mathbf{x}, \mathbf{y}) = \langle s(\mathbf{x})s(\mathbf{y}) \rangle = [\text{Prob}(++) + \text{Prob}(--)] - [\text{Prob}(+-) + \text{Prob}(-+)]$$

$$= [1 - \rho] - [\rho] = 1 - 2\rho. \tag{8.13}$$

Thus the interface density and the near-neighbor correlation function are related by $\rho = [1 - G(\mathbf{x}, \mathbf{y})]/2$. Using Eq. (8.12) for the correlation function, the time dependence of the interfacial density is then

$$\rho(t) \sim \begin{cases} t^{-1/2}, & d = 1, \\ (\ln t)^{-1}, & d = 2, \\ \mathcal{O}(1), & d > 2. \end{cases} \tag{8.14}$$

When $d \leq 2$, the probability that two voters at a given separation have opposite opinions asymptotically vanishes, and a coarsening mosaic of single-opinion domains develops. At the marginal dimension of $d = 2$, the coarsening process is slow and the density of interfaces asymptotically vanishes only as $1/\ln t$ (Fig. 8.1). In higher dimensions, an infinite system reaches a dynamic frustrated state, where opposite-opinion voters coexist and continually evolve in such a way that the mean density of each type of voter remains fixed.

The second basic consequence that follows from the correlation function is the time T_N to reach consensus for a finite system of N voters. To estimate the consensus time, we make use of the diffusive spread of the influence of any voter so that its influence range is of the order of $\sqrt{Dt}$. We then posit that consensus occurs when the total amount of correlation

within a distance of $\sqrt{Dt}$ of any voter equals the total number of voters N. This consensus criterion is

$$\int^{\sqrt{Dt}} G(r)\, r^{d-1}\, dr = N.$$

The lower limit can be set to 0 for $d = 1$ and should be set to the cutoff a for $d \geq 2$. Substituting the expressions for the correlation function given in Eq. (8.12) into this integral, the time dependence can be extracted by scaling and the asymptotic behavior is:

$$T_N \sim \begin{cases} N^2, & d = 1, \\ N \ln N, & d = 2, \\ N, & d > 2. \end{cases} \tag{8.15}$$

As the dimension decreases below 2, consensus takes progressively longer to achieve. This feature reflects the increasing difficulty in transmitting information when the dimensionality decreases.

Exact analysis

On a hypercubic d-dimensional lattice, the evolution of the pair correlation function is governed by the discrete diffusion equation (compare with Eq. (8.5))

$$\frac{dG(\mathbf{x}, t)}{dt} = -2G(\mathbf{x}, t) + \frac{1}{d} \sum_{\mathbf{y} \in \langle \mathbf{x} \rangle} G(\mathbf{y}, t), \tag{8.16}$$

subject to the uncorrelated initial condition $G(\mathbf{x}, 0) = \delta(\mathbf{x})$ and the boundary condition $G(\mathbf{0}, t) = 1$. Without the boundary condition, the solution is $I_{\mathbf{x}}(2t/d)e^{-2t}$ from Eq. (8.8). This solution agrees with the boundary condition only when $t = 0$. To maintain $G(\mathbf{0}, t) = 1$ throughout the evolution, we consider an initial value problem with a localized source at the origin; the strength of the source will be chosen to fulfill the boundary condition.

Since the governing equation is linear, the general solution is the linear combination

$$G(\mathbf{x}, t) = I_{\mathbf{x}}(2t/d)e^{-2t} + \int_0^t d\tau\, J(t - \tau)\, I_{\mathbf{x}}(2\tau/d)\, e^{-2\tau}. \tag{8.17}$$

The integral gives the contribution due to the effective time-dependent source $J(\tau)\, d\tau$ during the time interval $(\tau, \tau + d\tau)$. At the origin, Eq. (8.17) becomes

$$1 = \left[I_0(2t/d)\, e^{-2t/d} \right]^d + \int_0^t d\tau\, J(t - \tau) \left[I_0(2\tau/d)\, e^{-2\tau/d} \right]^d. \tag{8.18}$$

We now solve this integral equation to determine the strength of the source $J(t)$. The convolution structure of the integral in (8.18) suggests employing the Laplace transform. In terms of the Laplace transforms,

$$J(\sigma) = \int_0^\infty dt\, e^{-\sigma t}\, J(t) \quad \text{and} \quad \mathcal{I}_d(\sigma) = \int_0^\infty dt\, e^{-\sigma t} \left[I_0(2t/d)\, e^{-2t/d} \right]^d, \tag{8.19}$$

we deduce the simple relation

$$J(\sigma) = [\sigma \mathcal{I}_d(\sigma)]^{-1} - 1 . \tag{8.20}$$

Using the integral representation of the Bessel function, $I_0(\tau) = (2\pi)^{-1} \int_0^{2\pi} e^{\tau \cos q} \, dq$, we rewrite the latter transform in (8.19) as the integral

$$\mathcal{I}_d(\sigma) = \int_0^{2\pi} \cdots \int_0^{2\pi} \prod_{i=1}^{d} \frac{dq_i}{2\pi} \, \frac{1}{\sigma + \frac{2}{d} \sum_{1 \le i \le d} (1 - \cos q_i)} . \tag{8.21}$$

Equations (8.20) and (8.21) give the exact expression for the Laplace transform of the source strength.

The large-time behavior is ultimately related to the small-σ behavior of the corresponding Laplace transform. Using (8.21), we obtain the leading behavior of $\mathcal{I}_d(\sigma)$:

$$\mathcal{I}_d(\sigma) \simeq \begin{cases} 2^{-1} \sigma^{-1/2} & d = 1, \\ (2\pi)^{-1} \ln(1/\sigma) & d = 2, \\ \mathcal{I}_d(0) & d > 2. \end{cases} \tag{8.22}$$

Thus when $\sigma \to 0$, the integral (8.21) diverges if $d \le 2$, and it remains finite if $d > 2$; in particular, in three dimensions the explicit expression for the constant is given by

$$\mathcal{I}_3(0) = \int_0^{2\pi} \int_0^{2\pi} \int_0^{2\pi} \frac{dq_1 \, dq_2 \, dq_3}{(2\pi)^3} \, \frac{1}{2 - \frac{2}{3}(\cos q_1 + \cos q_2 + \cos q_3)} ,$$

$$= \frac{\sqrt{24}}{\pi^2} \Gamma\left(\frac{1}{24}\right) \Gamma\left(\frac{5}{24}\right) \Gamma\left(\frac{7}{24}\right) \Gamma\left(\frac{11}{24}\right) .$$

This type of integral is known as a *Watson integral* and it appears ubiquitously in lattice Green's function problems. Combining (8.20) and (8.22) the leading $\sigma \to 0$ behavior of the Laplace transform of the strength of the source is therefore

$$J(\sigma) \simeq \begin{cases} 2 \, \sigma^{-1/2}, & d = 1, \\ 2\pi \, \sigma^{-1} [\ln(1/\sigma)]^{-1}, & d = 2, \\ [\mathcal{I}_d(0)]^{-1} \sigma^{-1}, & d > 2. \end{cases} \tag{8.23}$$

Converting to the time domain, we obtain the long-time asymptotics of the source strength:

$$J(t) \simeq \begin{cases} 2 \, \pi^{-1/2} \, t^{-1/2}, & d = 1, \\ 2\pi \, (\ln t)^{-1}, & d = 2, \\ (\mathcal{I}_d(0))^{-1}, & d > 2. \end{cases} \tag{8.24}$$

Finally, to extract the asymptotic behavior of the density of reactive interfaces we use (8.17) and (8.18) to write[8]

$$\rho(t) = \frac{1}{2} e^{-2t} \left[I_0(2t/d)\right]^{d-1} \left[I_0(2t/d) - I_1(2t/d)\right]$$

$$+ \frac{1}{2} \int_0^t d\tau J(t - \tau) \left[I_0(2\tau/d)\right]^{d-1} \left[I_0(2\tau/d) - I_1(2\tau/d)\right] e^{-2\tau} . \qquad (8.25)$$

Combining this result with (8.24), the density of reactive interfaces behaves as

$$\rho(t) \simeq \begin{cases} (4\pi t)^{-1/2}, & d = 1, \\ \pi/(2 \ln t), & d = 2, \\ \text{const.}, & d > 2, \end{cases} \qquad (8.26)$$

in the long-time limit. These asymptotically exact results confirm the predictions of the continuum approach, Eq. (8.14).

8.3 Ising–Glauber model

The Ising model is the most widely studied model in statistical mechanics because of its simplicity, its utility for understanding phase transitions, and its surprisingly wide range of applications. In the classic Ising model, two-state spins exist on the sites of a regular lattice. The spins may take one of two values: $s(\mathbf{x}) = \pm 1$. The Hamiltonian for the system is

$$\mathcal{H} = -\sum_{\langle i,j \rangle} J s_i s_j , \qquad (8.27)$$

where the sum is over all distinct nearest-neighbor spin pairs $\langle i, j \rangle$. Thus every parallel pair of neighboring spins contributes $-J$ to the energy and every antiparallel pair contributes $+J$. All thermodynamic properties of the Ising model can be obtained from the partition function $Z = \sum \exp(-\beta \mathcal{H})$, where the sum is over all spin configurations of the system, and $\beta = 1/T$ is the inverse temperature.[9] For $J > 0$, ferromagnetism arises spontaneously, i.e. at zero external magnetic field, when the temperature T is less than a critical temperature T_c and the spatial dimension $d \geq 2$. In this ordered state, the magnetization is non-zero and distant spins are correlated. Above T_c, spins are spatially disordered, with equal numbers in the $+1$ and -1 states so that the magnetization is zero. Moreover, spatial correlations between spins decay exponentially with their separation.

While the equilibrium properties of the Ising model follow from the partition function, non-equilibrium properties depend on the nature of the spin dynamics. There is considerable freedom in formulating the dynamics that is constrained only by the detailed balance

---

[8] For concreteness, we consider interfaces perpendicular to one specified axis; the total density of interfaces is a factor d times larger. Thus here we define $\rho(t) = \frac{1}{2}[1 - G(1_x, t)]$, where 1_x refers to the lattice site $(1, 0, 0, \ldots)$.

[9] As mentioned previously, we always measure the temperature in energy units; equivalently the Boltzmann constant k_B is set to 1.
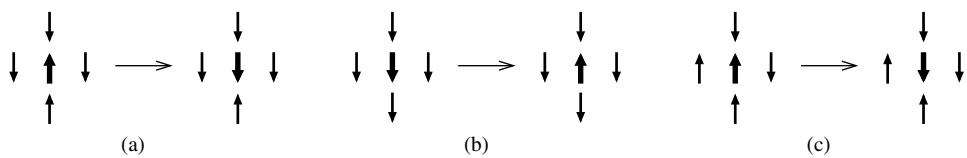
Fig. 8.4. (a) Energy-lowering, (b) energy-raising, and (c) energy-conserving spin-flip events for the Ising–Glauber model on the square lattice.

condition (see below and also the highlight on page 186). For example, spins may change one at a time or in correlated blocks, and the dynamics may or may not conserve the magnetization; magnetization-conserving dynamics will be treated in Section 8.7. This lack of uniqueness of the dynamical rules abounds in non-equilibrium statistical physics and it reflects the absence of universal principles for the evolution of non-equilibrium systems, such as free energy minimization in equilibrium statistical mechanics.

Transition rates

In the Glauber model, spins are selected one at a time in random order and each spin changes its state, $s_j \to -s_j$, at a rate that depends on its energy change in an update. By construction, the magnetization is *not* conserved in each update event. There are three types of transitions when a spin flips: energy-raising, energy-lowering, and energy-neutral (Fig. 8.4). Energy-raising events occur when a spin is aligned with a majority of its neighbors and energy-lowering events occur when a spin is anti-aligned with its local majority. Energy-neutral or energy-conserving events occur when the local magnetization is zero. The basic principle to fix the rates of these events is the *detailed balance condition*

$$P_{\rm eq}(\mathbf{s})\, w_j(\mathbf{s}) = P_{\rm eq}(\mathbf{s}^j)\, w_j(\mathbf{s}^j). \tag{8.28}$$

Here $\mathbf{s}$ denotes any of the 2^N states of all N spins in the system, $\mathbf{s}^j$ is the state derived from $\mathbf{s}$ in which the spin at j is flipped, $w_j(\mathbf{s})$ is the transition rate from $\mathbf{s}$ to $\mathbf{s}^j$, and $P_{\rm eq}(\mathbf{s})$ is the equilibrium Boltzmann weight

$$P_{\rm eq}(\mathbf{s}) = \frac{\exp[-\beta\, \mathcal{H}(\mathbf{s})]}{Z}\ . \tag{8.29}$$

Detailed balance is the statement of current conservation across each link in state space: the probability currents from $\mathbf{s}$ to $\mathbf{s}^j$ and from $\mathbf{s}^j$ to $\mathbf{s}$ – the left and right sides of Eq. (8.28) – are equal. By definition, the equilibrium weights satisfy the detailed balance condition. With transition rates defined by Eq. (8.28), any initial spin state will relax to the equilibrium state for all positive temperatures.

Detailed balance is the key principle that suggests how to implement a dynamics. However, several caveats need to be kept in mind.

1. The rates are *defined* by the detailed balance condition, but they are *not derived* from any microscopic model. Equilibrium statistical physics tells us that the equilibrium

Boltzmann weights (8.29) should be a solution of the non-equilibrium dynamics. This makes the choice of rates (8.28) natural for the system that starts at equilibrium. However, if the starting configuration is far from equilibrium, other rates can drive the system toward equilibrium.

2. The rates (8.28) ensure that equilibrium will be reached for any positive temperature because there is a non-zero probability current through each bond in the state space. However, when $T = 0$, energy-raising flips are forbidden and the corresponding transition rates are zero. Consequently, an Ising system can fall into a local minimum, that is, a state whose energy is lower than any other state that differs by one spin flip. There is no escape from such a metastable state and the system cannot reach the ground state.

Remarkably, the detailed balance condition can be satisfied with a simple choice for the transition rate. Consider the general inhomogeneous Ising model with Hamiltonian

$$\mathcal{H} = -\sum_{i,j} J_{ij} s_i s_j \,. \tag{8.30}$$

The couplings can be arbitrary (ferromagnetic or antiferromagnetic), the strengths can vary from bond to bond, and the number of neighbors can vary from spin to spin. Equation (8.28) gives

$$\frac{w_i(\mathbf{s})}{w_i(\mathbf{s}^i)} = \frac{P_{\mathrm{eq}}(\mathbf{s}^i)}{P_{\mathrm{eq}}(\mathbf{s})} = \frac{e^{-\beta s_i \sum J_{ij} s_j}}{e^{\beta s_i \sum J_{ij} s_j}} = \frac{1 - s_i \tanh\left(\beta \sum_{j\in\langle i\rangle} J_{ij} s_j\right)}{1 + s_i \tanh\left(\beta \sum_{j\in\langle i\rangle} J_{ij} s_j\right)} \,, \tag{8.31}$$

where the sum is over the nearest neighbors $\langle i \rangle$ of site i. In deriving the last equality we use a useful identity (valid for arbitrary $s = \pm 1$)

$$e^{As} = \cosh A + s \sinh A = \cosh A \left(1 + s \tanh A\right) . \tag{8.32}$$

The simplest set of rates that satisfy (8.31) is

$$w_i(\mathbf{s}) = \frac{1}{2}\left[1 - s_i \tanh\left(\beta \sum_{j\in\langle i\rangle} J_{ij} s_j\right)\right] , \tag{8.33}$$

where we set the transition rate for energy-conserving flips to $1/2$.[10] Before turning to this situation, let's consider an even simpler version of Glauber dynamics to illustrate its workings – the mean-field limit.

---

[10] In the one-dimensional case, the rate (8.33) was proposed by Glauber.

8.4 Mean-field approximation

The term *mean-field approximation* refers to an effective field that represents a substitute for the local interactions between spins. The starting point of the mean-field approximation involves no approximation; we just rewrite the Hamiltonian (8.27) in the form

$$\mathcal{H} = -\sum_i h_i s_i, \qquad h_i = \frac{1}{2} \sum_{j \in \langle i \rangle} J s_j, \tag{8.34}$$

and view the Hamiltonian as that of a collection of spins $\{s_i\}$ that are each influenced by a local inhomogeneous magnetic field h_i that originates from the neighboring spins $j \in \langle i \rangle$.[11] The approximation comes when we replace the spin s_j at each site $j \in \langle i \rangle$ by its mean value – the average magnetization m – so that the local field h_i reduces to the homogeneous "mean field" $h = Jzm/2$, where z is the lattice coordination number. This replacement maps the initial interacting system into a system of effectively independent spins that interact only with an external field h and not directly with their neighbors.

The lack of a direct interaction between the neighbors leads to a crucial property: neighboring spins are uncorrelated. Mathematically

$$\langle s_i s_j \rangle = \langle s_i \rangle \langle s_j \rangle \quad \text{for} \quad i \neq j, \tag{8.35}$$

and similarly for higher-order correlation functions. In many situations, particularly in non-equilibrium processes that do not admit a Hamiltonian description, it is not possible to introduce a "mean field" that couples to all individual spins (or particles) in the system. Such systems are naturally described in terms of correlation functions. However, the governing equations for these correlation functions are usually hierarchical; that is, the n-body correlation function is coupled to the $(n+1)$-body correlation function. In this case, factorizability of the correlation functions, as in (8.35), *defines* a mean-field approximation. This approximation replaces the exact hierarchical equations for the correlation functions by mean-field equations that are recursive and closed, and therefore tractable. Generally, factorizability of the correlation functions represents an uncontrolled approximation, but it often captures essential qualitative features of interacting systems.

By construction, the mean-field description cannot be quantitatively accurate for systems with local interactions, but the mean-field description is exact in the special case where all pairs of elements interact. A particularly simple example is the ferromagnetic Ising model on a complete graph of N sites[12] where the exact solution coincides with the mean-field solution. The Hamiltonian of the system is

$$\mathcal{H} = -\frac{1}{N} \sum_{i<j} s_i s_j . \tag{8.36}$$

---

[11] The factor $1/2$ in h_i corrects for double counting, since each bond is counted twice in the Hamiltonian in Eq. (8.34).

[12] In a complete graph of N sites, there are $N(N-1)/2$ bonds since all sites are connected.

Here, the interaction strength is chosen to scale inversely with the system size so that the energy is extensive, i.e. the energy scales linearly with N. Since each spin interacts with all other spins, the local field is the same as the magnetization.[13]

We now study Glauber dynamics on the complete graph, for which the transition rate, Eq. (8.33), becomes

$$w_i = \frac{1}{2}\left[1 - s_i \tanh\left(\frac{\beta}{N}\sum_j{}' s_j\right)\right].$$ (8.37)

The prime on the sum denotes that it is over all spins in the system except for s_i. When $N \gg 1$, we can include the s_i term with negligible error. We can also replace the exact magnetization $N^{-1}\sum_j s_j$ by its average value $m = N^{-1}\sum_j S_j$, where $S_j = \langle s_j \rangle$. The equation of motion for the average spin $dS_i/dt = -2\langle s_i w_i \rangle$ becomes

$$\frac{dS_i}{dt} = -S_i + \tanh\beta m.$$ (8.38)

Summing Eqs (8.38) over all sites, the average magnetization satisfies

$$\frac{dm}{dt} = -m + \tanh\beta m.$$ (8.39)

In contrast to one dimension, the magnetization is generally not conserved. In equilibrium, Eq. (8.39) predicts that a phase transition occurs at $\beta_c = 1$. When $\beta \leq \beta_c = 1$, the only solution of $m = \tanh\beta m$ is $m = 0$. For $\beta > \beta_c$, there are two non-trivial equilibrium solutions, $m = \pm m_{eq}$, where m_{eq} is the positive root of the transcendental equation $m = \tanh(\beta J m)$, in addition to the zero-magnetization solution. By a linear stability analysis, it is easy to show that the states with $m = \pm m_{eq}$ are stable for $\beta > 1$, while the zero-magnetization state is unstable. Near the critical temperature $m_{eq} \simeq \sqrt{3(\beta - 1)} = \sqrt{3(T_c - T)/T}$ for $T \lesssim T_c = 1$. The emergence of two equivalent, but symmetry-breaking, ground states when the Hamiltonian itself is symmetric is known as *spontaneous symmetry breaking*.

Let's return now to the dynamics. Above the critical temperature, we expand $\tanh\beta m$ in Eq. (8.39) in powers of βm to give

$$\frac{dm}{dt} = -(\beta_c - \beta)m - \frac{1}{3}(\beta m)^3 + \mathcal{O}(m^5).$$ (8.40)

In the high-temperature phase, the cubic and higher-order terms are negligible, and the magnetization decays exponentially with time, $m \sim e^{-t/\tau}$, with $\tau = (\beta_c - \beta)^{-1}$. At the critical point, $dm/dt = -m^3/3 + \mathcal{O}(m^5)$, so that the magnetization has the power-law relaxation

$$m \sim t^{-1/2}.$$ (8.41)

---

[13] There is an alternative network – the Cayley tree – where an exact analysis is possible, that also leads to the same critical behavior as the complete graph. These two examples represent opposite topological extremes: the complete graph contains the maximum possible number of closed loops, while the Cayley tree is a regular graph without any loops.

Below the critical temperature, the $m = 0$ solution is unstable and the magnetization approaches $\pm m_{\text{eq}}$. Expanding Eq. (8.40) about the equilibrium magnetization, we find that the magnetization decays to its equilibrium value exponentially in time, $|m - m_{\text{eq}}| \sim e^{-t/\tau}$, with $\tau^{-1} = 1 - \beta / \cosh^2(\beta m_{\text{eq}})$. Thus as the critical point is approached, either from above or from below, the relaxation time scale diverges as

$$\tau \sim |T_c - T|^{-1}. \tag{8.42}$$

The divergence of the relaxation time as $T \to T_c$ is a sign of critical slowing down in the approach to equilibrium.

   These distinct behaviors for supercritical, critical, and subcritical dynamics reflect the central dogma outlined at the beginning of this chapter. Above the critical temperature, the system quickly relaxes to an equilibrium $m = 0$ disordered state. Spin correlations are short-ranged for all finite $T > T_c$ and are absent for $T = \infty$. At the critical temperature, the system relaxes much more slowly. For all temperatures below the critical temperature, the system relaxes to one of the equilibrium ordered states. Because of this universal behavior, we focus on the dynamics at zero temperature that represents subcritical dynamics in its cleanest form.

8.5 Glauber dynamics in one dimension

In one dimension, the flip rate (8.33) becomes

$$w_i(\mathbf{s}) = \frac{1}{2}\left[1 - s_i \tanh\left(2\beta J \frac{s_{i-1} + s_{i+1}}{2}\right)\right] = \frac{1}{2}\left[1 - \gamma s_i \frac{s_{i-1} + s_{i+1}}{2}\right],$$
$$\gamma = \tanh(2\beta J), \tag{8.43}$$

where in the last step we use the identity $\tanh \epsilon x = \epsilon \tanh x$ when $\epsilon = 0, \pm 1$, as in the Glauber flip rate. While this rate is compatible with detailed balance, it is just one of many such compatible rates. To gain a deeper appreciation of Glauber dynamics, let us seek the most general spin-flip rate that satisfies the following natural requirements:

1. *Locality*. Since the Hamiltonian $\mathcal{H} = -J \sum s_j s_{j+1}$ involves only nearest-neighbor interactions, the spin-flip rates should also depend only on the nearest neighbors of each spin: $w_i(\mathbf{s}) = w_i(s_{i-1}, s_i, s_{i+1})$.
2. *Left/right symmetry*. Invariance under the interchange $i + 1 \leftrightarrow i - 1$.
3. *Up/down symmetry*. Invariance under reversal of all spins to reflect the invariance of the Hamiltonian under this transformation. In a non-zero magnetic field, up/down symmetry no longer applies and flip rates are not invariant under spin reversal.

The most general flip rate that satisfies these conditions has the form

$$w_i(s_{i-1}, s_i, s_{i+1}) = A + B s_i(s_{i-1} + s_{i+1}) + C s_{i-1} s_{i+1}. \tag{8.44}$$

An instructive way to establish this result is to expand the spin-flip rate in a Taylor series as

$$w_i(s_{i-1}, s_i, s_{i+1}) = \sum_{a,b,c \geq 0} A_{abc}\, s_{i-1}^a s_i^b s_{i+1}^c \,.$$

Because $s^2 = 1$, the infinite series reduces to a finite sum in which a, b, c equal 0 or 1. Invoking left/right and up/down symmetries reduces the sum to the rate in Eq. (8.44).

By redefining the time variable, $t \to 2At$, we can set $A = 1/2$. The resulting flip rate depends on only two parameters and it is traditionally written as

$$w_i = \frac{1}{2}\left[1 - \frac{\gamma}{2} s_i\,(s_{i-1} + s_{i+1}) + \delta s_{i-1} s_{i+1}\right]. \tag{8.45}$$

Pictorially, this flip rate is given by

$$w_i = \frac{1}{2} \times
\begin{cases}
1 - \gamma + \delta, & \text{for spin states } \uparrow\uparrow\uparrow, \ \downarrow\downarrow\downarrow, \\
1 - \delta, & \text{for spin states } \uparrow\uparrow\downarrow, \ \downarrow\uparrow\uparrow, \ \downarrow\downarrow\uparrow, \ \uparrow\downarrow\downarrow, \\
1 + \gamma + \delta, & \text{for spin states } \uparrow\downarrow\uparrow, \ \downarrow\uparrow\downarrow,
\end{cases} \tag{8.46}$$

where the constants γ and δ are constrained by the detailed balance condition

$$\frac{1 - \frac{1}{2}\gamma s_j(s_{j-1} + s_{j+1}) + \delta s_{i-1} s_{i+1}}{1 + \frac{1}{2}\gamma s_j(s_{j-1} + s_{j+1}) + \delta s_{i-1} s_{i+1}} = \frac{1 - \frac{1}{2}\tanh(2\beta J) s_j(s_{j-1} + s_{j+1})}{1 + \frac{1}{2}\tanh(2\beta J) s_j(s_{j-1} + s_{j+1})}\,. \tag{8.47}$$

By evaluating this equality for all configurations of the spin triplets (s_{i-1}, s_i, s_{i+1}), detailed balance is satisfied when

$$\gamma = (1 + \delta)\,\tanh(2\beta J)\,, \tag{8.48}$$

and the physical requirement that the flip rates are non-negative provides the bound $-1 \leq \delta \leq 1$.

While the flip rate (8.45) is the most general, the Glauber rate (8.43) is the most natural. For example, in the $T \to \infty$ limit the Glauber rate predicts that all flip events are equiprobable, while the general rate does not. Moreover, the kinetic Ising model in one dimension is solvable only for the Glauber rate (as well as for an exceptional case that we do not discuss). In the following, we always use the Glauber spin-flip rate unless stated otherwise.

Magnetization and correlation functions

Following the same line of analysis as that given for the voter model (Eqs (8.3) and (8.5)), the equations of motion for the low-order correlation functions are

$$\frac{dS_j}{dt} = -2\langle s_j w_j\rangle, \tag{8.49a}$$

$$\frac{dS_{i,j}}{dt} = -2\langle s_i s_j [w_i + w_j]\rangle, \tag{8.49b}$$

where $S_i = \langle s_i\rangle$ and $S_{i,j} = \langle s_i s_j\rangle$.

Using the transition rate (8.43) and the identity $s_j^2 = 1$, the evolution equation for the average spin S_j is

$$\frac{dS_j}{dt} = -S_j + \frac{\gamma}{2}\left(S_{j-1} + S_{j+1}\right). \tag{8.50}$$

With the initial condition $S_j(0) = \delta_{j,0}$, the solution is (see the highlight on page 14 on Bessel functions)

$$S_j(t) = I_j(\gamma t)\, e^{-t}, \tag{8.51}$$

where I_j is the modified Bessel function of order j. Notice that at $T = 0$, corresponding to $\gamma = 1$, the equation of motion (8.50) is identical to that of the one-dimensional voter model. In fact, the two models are identical in one dimension for $T = 0$. When $T > 0$, the new feature of the Ising–Glauber model compared to the voter model is the presence of the temperature-dependent factor γ. Consequently the average spin at any site asymptotically decays as $S_j(t) \sim (2\pi\gamma t)^{-1/2} e^{-(1-\gamma)t}$; namely, exponential decay with relaxation time $\tau = (1-\gamma)^{-1}$, and algebraic decay for $T = 0$. The magnetization $m = N^{-1}\sum_j S_j$ satisfies $dm/dt = -(1-\gamma)m$, so that m decays exponentially with time at any positive temperature,

$$m(t) = m(0)\, e^{-(1-\gamma)t}, \tag{8.52}$$

and is conserved only at zero temperature, just as in the voter model. The one-dimensional Ising–Glauber model undergoes critical slowing down – slower relaxation at the critical point ($T = 0$ in one dimension) than for $T > 0$. For a general initial condition, the general solution to Eq. (8.50) is a linear combination of the partial solutions (8.51), and thus has the form

$$S_\ell(t) = \sum_{m=-\infty}^{\infty} \sigma_m\, I_{\ell-m}(\gamma t)\, e^{-t}. \tag{8.53}$$

Let's now study the pair correlation function, $S_{i,j}$. The nearest-neighbor correlation function is particularly useful because it has a simple geometrical interpretation in terms of domain walls (Fig. 8.5). Two neighboring antiparallel spins correspond to an intervening domain-wall quasi-particle, while two neighboring parallel spins correspond to no domain

(a)   (b)   (c)
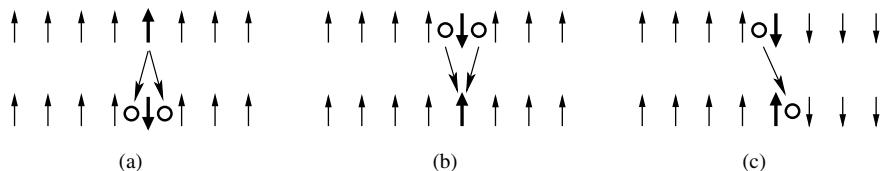
Fig. 8.5.   Mapping between states of the Ising or the voter models in one dimension and corresponding domain-wall quasi-particles: (a) an energy-raising move and creation of a pair of domain walls; (b) an energy-lowering move and annihilation of domain walls; and (c) an energy-conserving move and diffusion of a domain wall.

wall. That is, the bond $(k, k+1)$ hosts a domain wall if $\frac{1}{2}(1 - s_k s_{k+1}) = 1$. The density of domain walls is given by

$$\rho = \left\langle \tfrac{1}{2}(1 - s_k s_{k+1}) \right\rangle = \tfrac{1}{2}(1 - G_1).$$

For translationally invariant systems, the correlation function depends only the separation of the two spins, $G_k \equiv S_{i,i+k}$, so that the master equation (8.49b) becomes

$$\frac{dG_k}{dt} = -2G_k + \gamma \left(G_{k-1} + G_{k+1} \right) \tag{8.54}$$

for $k > 0$, subject to the boundary condition $G_0(t) = \langle s_i^2 \rangle = 1$. To obtain the equilibrium solution to Eq. (8.54), we substitute in $G_k(\infty) \sim \eta^k$ and find that the right-hand side vanishes when $2 = \gamma(\eta + \eta^{-1})$. From this relation, we deduce that $\eta = [1 - \sqrt{1 - \gamma^2}]/\gamma = \tanh \beta J$. Thus the equilibrium solution can be written as

$$G_k(\infty) = e^{-k/\xi}, \qquad \text{with} \qquad \xi = [\ln(\coth \beta J)]^{-1}, \tag{8.55}$$

in agreement with equilibrium statistical physics. As expected, the correlation length ξ diverges as $T \to 0$, indicative of a phase transition, and ξ vanishes at infinite temperature.

To find the time dependence of the correlation function, note that the evolution equation (8.54) for the correlation function has the same form as that for the mean spin, apart from an overall factor of 2 and the fixed boundary condition at the origin. Thus the general solution must be built from components of the same form as (8.53), with the replacement of $t \to 2t$. We now need to determine the appropriate linear combination of these component solutions that simultaneously satisfies the initial condition $G_k(t = 0) = \delta_{k,0}$ and the boundary condition $G_0 = 1$. One piece of the full solution is just the equilibrium correlation function $G_k(\infty) = \eta^k$. To this, we must add the general homogeneous solution that satisfies the prescribed constraints. Pictorially, the appropriate initial condition for the homogeneous solution consists of an arbitrary odd function plus an antisymmetric piece that cancels the equilibrium solution for $k > 0$. The antisymmetry of these pieces ensures that $G_0 = 1$ and that the prescribed initial condition is satisfied for $k > 0$. The general solution for $k > 0$ therefore is

$$
\begin{aligned}
G_k(t) &= \eta^k + e^{-2t} \sum_{\ell=-\infty}^{\infty} G_\ell(0) I_{k-\ell}(2\gamma t) \\
&= \eta^k + e^{-2t} \sum_{\ell \geq 1} [G_\ell(0) - \eta^\ell] I_{k-\ell}(2\gamma t) + e^{-2t} \sum_{\ell \leq -1} [G_\ell(0) + \eta^{|\ell|}] I_{k-\ell}(2\gamma t) \\
&= \eta^k + e^{-2t} \sum_{\ell \geq 1} [G_\ell(0) - \eta^\ell][I_{k-\ell}(2\gamma t) - I_{k+\ell}(2\gamma t)]. \tag{8.56}
\end{aligned}
$$

The most interesting behavior arises when the temperature equals zero and we henceforth assume $T = 0$ unless stated otherwise. For the random initial condition with initial magnetization $m_0 = 0$, we have $G_\ell(0) = 0$, and the time dependence of the correlation

function is given by

$$G_k(t) = 1 - e^{-2t} \left[I_0(2t) + I_k(2t) + 2 \sum_{j=1}^{k-1} I_j(2t) \right]. \tag{8.57}$$

From this expression, the domain-wall density is simply

$$\rho(t) = \frac{1}{2}(1 - G_1) = \frac{1}{2} e^{-2t} \left[I_0(2t) + I_1(2t) \right] \simeq (4\pi t)^{-1/2}, \tag{8.58}$$

Another illustrative example is the antiferromagnetic initial condition, where $G_\ell(0) = (-1)^\ell$. Now the domain-wall density is

$$\rho(t) = e^{-2t} I_0(2t) \simeq (4\pi t)^{-1/2}, \tag{8.59}$$

which has the same asymptotic decay as the random initial condition. Generally, the domain-wall density depends only on the initial magnetization; other details of the initial condition are irrelevant. Finally, note that, at zero temperature, the domain-wall dynamics coincides with *irreversible diffusion-controlled annihilation*, in which a population of identical particles diffuses freely and two particles annihilate whenever they meet. Similarly, for $T > 0$, Ising–Glauber dynamics is equivalent to diffusion-controlled annihilation with pair creation. We will return to this subject in Chapter 13.

Example 8.1. *Domain-wall multiplets.* We can extend the analysis of domain walls and determine the probability that two adjacent bonds contain domain walls. Bonds $(k-1, k)$ and $(k, k+1)$ host domain walls if $\frac{1}{4}(1 - s_{k-1}s_k) \times (1 - s_k s_{k+1}) = 1$. Therefore the density of domain-wall doublets is

$$\rho_2 = \left\langle \frac{1}{4}(1 - s_{k-1}s_k) \times (1 - s_k s_{k+1}) \right\rangle = \frac{1}{4}(1 - 2G_1 + G_2).$$

Substituting in G_1 and G_2 from (8.57), we find that for the uncorrelated initial condition with $m_0 = 0$ the doublet density is

$$\rho_2(t) = \frac{1}{4} e^{-2t} \left[I_0(2t) - I_2(2t) \right] = \frac{1}{4t} e^{-2t} I_1(2t) \simeq \frac{1}{8\sqrt{\pi t^3}} \qquad \text{as} \quad t \to \infty. \tag{8.60}$$

Similarly, the triplet density

$$\rho_3 = \left\langle \frac{1}{8}(1 - s_{k-1}s_k) \times (1 - s_k s_{k+1}) \times (1 - s_{k+1}s_{k+2}) \right\rangle$$

can be expressed as

$$\rho_3 = \frac{1}{4}(1 - 3G_1 + 2G_2 - G_3 + \langle s_{k-1}s_k s_{k+1}s_{k+2} \rangle). \tag{8.61}$$

Thus to determine the triplet density we need to compute the four-spin correlation function $\langle s_{k-1}s_k s_{k+1}s_{k+2} \rangle$. This is not easy, but for the most natural uncorrelated initial condition with $m_0 = 0$ the answer is relatively simple (problem 8.5):

$$\langle s_{k-1}s_k s_{k+1}s_{k+2} \rangle = G_1^2 + G_1 G_3 - G_2^2. \tag{8.62}$$

Combining this result with (8.61) and using G_1, G_2, and G_3 from (8.57) we obtain

$$\rho_3 = \frac{e^{-4t}}{4} \left[I_0^2(2t) + I_0(2t)I_1(2t) + I_0(2t)I_3(2t) - I_1^2(2t) \right.$$
$$\left. -2I_1(2t)I_2(2t) + I_1(2t)I_3(2t) - I_2^2(2t) \right]. \tag{8.63}$$

Asymptotically, $\rho_3 \simeq 1/(8\pi t^3)$.

In the presence of a non-zero magnetic field, the Hamiltonian of the ferromagnetic Ising spin chain is

$$\mathcal{H} = -J \sum_j s_j s_{j+1} - H \sum_j s_j, \tag{8.64}$$

and one should expect that spins will align with the field. However, the introduction of a field leads to an intractable evolution equation for the kinetic Ising model at positive temperature. However, at zero temperature, the model is simple and soluble, and nicely illustrates the asymptotic alignment of spins with the field.

Example 8.2. *Ising chain in a magnetic field.* The detailed balance condition now gives

$$\frac{w_i(\mathbf{s})}{w_i(\mathbf{s}^i)} = \frac{e^{-\beta J s_i(s_{i-1}+s_{i+1})-\beta H s_i}}{e^{\beta J s_i(s_{i-1}+s_{i+1})+\beta H s_i}} = \frac{1 - \frac{1}{2}\gamma s_i (s_{i-1}+s_{i+1})}{1 + \frac{1}{2}\gamma s_i (s_{i-1}+s_{i+1})} \frac{1 - h s_i}{1 + h s_i}, \tag{8.65}$$

where $\gamma = \tanh 2\beta J$ and $h = \tanh \beta H$, and we take the field to be positive without loss of generality. A natural choice of the transition rate is therefore

$$w_i(\mathbf{s}) = \frac{1}{2}\left[1 - \frac{\gamma}{2}s_i(s_{i-1}+s_{i+1}) - h s_i + \frac{\gamma}{2}h(s_{i-1}+s_{i+1})\right]. \tag{8.66}$$

The magnetic field gives rise to terms linear in spin variables, which are of lower order than the quadratic field-independent term. This linearity suggests that the spin dynamics in a non-zero field could be solvable. Unfortunately, this is not the case. For example, the average spin satisfies

$$\frac{dS_i}{dt} = -S_i + \frac{\gamma}{2}(S_{i-1}+S_{i+1}) + h - \frac{\gamma h}{2}(S_{i-1,i}+S_{i,i+1}) \tag{8.67}$$

that couples to the pair correlation function of adjacent spins $S_{i-1,i}$. Similarly, the equations for the pair correlation function contain three-spin correlations, and this hierarchical pattern proceeds *ad infinitum*.

While the behavior at positive temperature is analytically intractable, the case of $T = 0$ is simpler and still interesting. To determine the transition rates at $T = 0$ it is expedient to use the over-arching rule that a spin-flip event that decreases the energy is allowed and all other events are forbidden. This rule leads to the conclusion that there are two different dynamics depending on whether $H > 2J$ or $H < 2J$.

Strong field ($H > 2J$). In this trivial case, up spins do not evolve, while down spins flip with rate 1 independent of the state of neighboring spins. Therefore the transition rate is $w_i = (1 - s_i)/2$ and the mean spin evolves according to $\dot{S}_i = 1 - S_i$. Starting from an uncorrelated initial condition with zero magnetization, the solution is $S_i = 1 - e^{-t}$, and the fully aligned state is reached exponentially quickly. Since spins are uncoupled, all correlation functions factorize, $S_{i_1,\ldots,i_n} = \left(1 - e^{-t}\right)^n$.

Weak field ($0 < H < 2J$). In this case, out of eight possible configurations of three consecutive spins, the central spin can flip with unit rate only in the following three cases:

$$\uparrow\downarrow\downarrow \Longrightarrow \uparrow\uparrow\downarrow, \qquad \uparrow\downarrow\uparrow \Longrightarrow \uparrow\uparrow\uparrow, \qquad \downarrow\uparrow\downarrow \Longrightarrow \downarrow\downarrow\downarrow .$$

In the domain picture, down-spin domains shrink systematically. The density P_n of such domains of length n evolves according to

$$\frac{dP_n}{dt} = 2(P_{n+1} - P_n) \tag{8.68}$$

when $n \geq 2$; a slightly different equation $dP_1/dt = 2P_2 - P_1$ describes the evolution of down-spin domains of length one. These equations are readily soluble (problem 8.9).
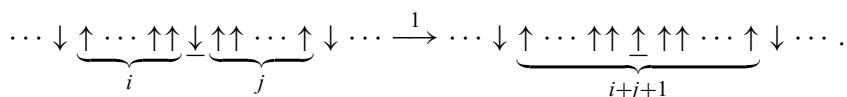
Domain length distribution

At zero temperature, domain walls that diffuse and annihilate upon meeting are the fundamental excitations of the Ising chain with Glauber dynamics. A basic measure of these excitations is their mean separation (equivalent to the inverse of the domain-wall density) which grows with time as $t^{1/2}$ at zero temperature. A more fundamental quantity is the *distribution* of distances, i.e. the probability P_k to find a domain of length k. We define this probability per unit length to give an intensive quantity. A domain consists of k consecutive aligned spins that is bounded by oppositely oriented spins at each end. We can obtain partial information about this distribution by noting that the domain-wall density ρ is simply $\sum_{k \geq 1} P_k$, which should scale as $t^{-1/2}$. Similarly, the normalization condition gives $\sum_{k \geq 1} k P_k = 1$. These facts suggest that the domain length distribution has the scaling form

$$P_k(t) \simeq C\, t^{-1}\, \Phi(k t^{-1/2}). \tag{8.69}$$

The normalization requirement then implies $\int_0^\infty x\, \Phi(x)\, dx = 1$, while equating $\sum_{k \geq 1} P_k$ with ρ and using the asymptotic decay in Eq. (8.58) gives $\int_0^\infty \Phi(x)\, dx = (4\pi)^{-1/2} \equiv C$.

We can also infer the short-distance tail of the scaling function $\Phi(x)$ from the long-time decay of the domain density. Consider the role of the shortest possible domain of length 1 in the equation for the domain density ρ. When a single-spin domain flips, three domains merge into a single larger domain, as illustrated by

$$\cdots \downarrow \underbrace{\uparrow \cdots \uparrow\uparrow}_{i} \downarrow \underbrace{\uparrow\uparrow \cdots \uparrow}_{j} \downarrow \cdots \xrightarrow{1} \cdots \downarrow \underbrace{\uparrow \cdots \uparrow\uparrow \uparrow \uparrow\uparrow \cdots \uparrow}_{i+j+1} \downarrow \cdots .$$

Since such events occur with unit rate, the domain density decays as

$$\frac{d\rho}{dt} = -2P_1 \, . \tag{8.70}$$

Using Eq. (8.58), we obtain $P_1 \simeq \frac{1}{4} C t^{-3/2}$. On the other hand, expanding the scaled distribution Φ in a Taylor series gives $P_1 = \Phi(0)t^{-1} + \Phi'(0) t^{-3/2} + \cdots$. Comparing these two results, we deduce that $\Phi(0) = 0$ and $\Phi'(0) = \frac{1}{4}C$. Therefore the scaling function vanishes linearly in the limit of small argument x:

$$\Phi(x) = \frac{1}{4} C x + \cdots \qquad \text{as} \qquad x \to 0. \tag{8.71}$$

The linearity of the small-length tail is a generic feature of many one-dimensional interacting many-body systems. This linearity arises because diffusive annihilation of domain walls is equivalent to the distance of a single random walk from an absorbing origin. This latter system generically has a linear dependence of the probability distribution near the origin.

  While scaling provides clues about the nature of the length distribution, the full solution for this distribution is not known. Here we present an approximate solution that is based on the *independent interval approximation* in which the lengths of neighboring domains are assumed to be uncorrelated. This approximation has been successfully applied to find approximate, but quite accurate, solutions to a variety of one-dimensional domain evolution and reaction processes. Within the independent interval approximation, the length distribution evolves according to

$$\frac{dP_k}{dt} = -2P_k + P_{k+1} + P_{k-1} \left(1 - \frac{P_1}{\rho}\right) + \frac{P_1}{\rho^2} \sum_{i+j=k-1} P_i P_j - \frac{P_1}{\rho} P_k. \tag{8.72}$$

The first three terms account for length changes due to a domain-wall hopping. The prefactor $(1-P_1/\rho)$ multiplying P_{k-1} ensures that the neighbor of a $(k-1)$-domain has length greater than 1, so that the domain growth process $k-1 \to k$ actually leads to $(k-1,j) \to (k,j-1)$, and not to $(k-1,1,j) \to (k+j)$. The last two terms account for changes in the length distribution due to domain mergers. Because any merger requires a domain of length 1, both these terms are proportional to P_1. The gain term accounts for the merger of three domains of lengths i, j, and 1, with $i+j+1 = k$. The loss term accounts for the merger of a domain of length k with a domain of any length. The independent interval approximation has been used to write the probability of finding either three (or two) adjacent domains of specified length as the product of single-domain densities. Equation (8.72) remains valid when $k = 1$ if we set $P_0 = 0$.

  Since the typical domain length grows continuously with time, we replace the sums by integrals and the integer k by the continuous variable x to convert (8.72) into a continuous equation. We now substitute the scaling form (8.69), as well as $\rho \simeq C t^{-1/2}$, and $P_1 \simeq \frac{1}{4} Ct^{-3/2}$, into this continuum version of the master equation and find that the scaling

function satisfies

$$\frac{d^2\Phi}{dx^2} + \frac{1}{2}\frac{d(x\Phi)}{dx} + \frac{1}{4C}\int_0^x \Phi(y)\,\Phi(x-y)\,dy = 0. \tag{8.73}$$

Because the last term is a convolution, we introduce the Laplace transform $\Phi(s) = (1/C)\int_0^\infty \Phi(x)\,e^{-sx}\,dx$ (albeit with a prefactor) to express the convolution as a product and recast Eq. (8.73) as the ordinary differential equation

$$\frac{d\Phi}{ds} = \frac{\Phi^2}{2s} + 2s\Phi - \frac{1}{2s}\,, \tag{8.74}$$

subject to the boundary condition $\Phi(s=0) = 1$, to satisfy normalization.

   Equation (8.74) is a Riccati equation. Occasionally it is possible to solve such an equation by converting it into a second-order linear differential equation that happens to be soluble. In the present case, the transformation $\Phi = -2s\,d\ln\psi/ds$ recasts (8.74) into

$$\frac{d^2\psi}{ds^2} + \frac{d\psi}{ds}\left(\frac{1}{s} - 2s\right) - \frac{\psi}{4s^2} = 0.$$

Now we apply the standard trick of eliminating the linear term to transform to the Schrödinger equation:

$$\frac{d^2y}{ds^2} + (2 - s^2)y = 0\,. \tag{8.75}$$

This elimination is achieved by seeking a solution in a form $\psi = yv$ and then forcing the term linear in y' to vanish. This condition fixes the integrating factor to be $v = \exp\left[-\frac{1}{2}\int (1/s - 2s)\,ds\right]$. The relation between the original variable and y is given by $\Phi = 1 - 2s^2 - 2s\,d\ln y/ds$. Equation (8.75) is the *parabolic cylinder equation*, which arises, for example, in the quantum harmonic oscillator. The general solution to this equation is a linear combination of two linearly independent terms, $y = C_+ D_{1/2}(\sqrt{2}\,s) + C_- D_{1/2}(-\sqrt{2}\,s)$, where $D_\nu(x)$ is the parabolic cylinder function of order ν. From the large-s behavior, $\Phi \simeq (4s)^{-2}$, together with the asymptotics of $D_\nu(s)$, it follows that $C_- = 0$. Therefore the Laplace transform is

$$\Phi = 1 - 2s^2 - 2s\frac{d}{ds}\ln D_{1/2}(\sqrt{2}\,s). \tag{8.76}$$

To determine the large-length tail of the domain length distribution, we investigate the small-s limit of the Laplace transform (8.76). In this limit, the parabolic cylinder function has a simple pole of the form $\Phi \simeq 2\lambda(s + \lambda)^{-1}$, with $\lambda \doteq 0.5409$ equal to the first zero of the parabolic cylinder function, $D_{1/2}(-\sqrt{2}\,\lambda) = 0$. Inverting the Laplace transform gives our main result that the large-x tail of $\Phi(x)$ decays exponentially with x, $\Phi(x) \simeq Ae^{-\lambda x}$.

   While the independent interval approximation is not exact, it leads to a tractable equation for the domain length distribution whose solution yields the main qualitative features of the domain length distribution including: (i) the linear small-length limit, (ii) the large-length exponential tail, and (iii) correct integrated properties, such as the $t^{-1/2}$ decay of the number of domains.

8.6 Glauber dynamics in higher dimensions
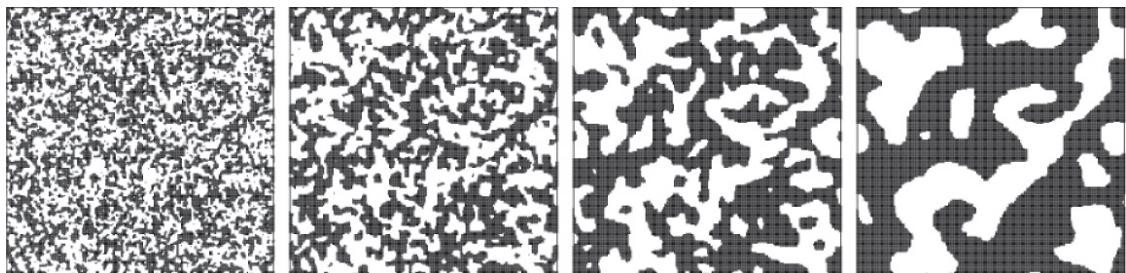
Spin evolution generically leads to intricate growing patterns, or domain coarsening (Fig. 8.6). While general understanding of the Ising–Glauber model in $d > 1$ is primarily based on continuum theories (Chapter 9), we can gain many insights from studying kinetic discrete-spin systems.

Transition rates
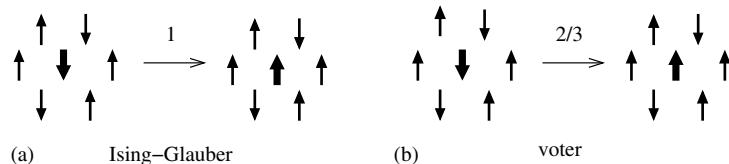
For Glauber dynamics in $d > 1$, the transition rate is

$$w_i(\mathbf{s}) = \frac{1}{2}\left[1 - s_i \tanh\left(\beta J \sum_{j\in\langle i\rangle} s_j\right)\right], \tag{8.77}$$

where the sum is over the $2d$ nearest neighbors of spin i on a d-dimensional hypercubic lattice. Let us first try to understand why the Ising–Glauber model at zero temperature is not soluble for $d > 1$, while the closely related voter model is soluble for all d. At zero temperature, the allowed spin-flip events and their corresponding rate define the *majority rule* – a spin flips to agree with the majority of its neighbors (Fig. 8.7). In contrast, as discussed in Section 8.2, the voter model obeys the *proportional rule* – a voter adopts the state of its local majority with probability equal to the fraction of neighbors in this state. This proportionality allows the master equation of the voter model in d dimensions to be

Fig. 8.6.   Evolution of the Ising–Glauber model at $T = 0$ on a 256×256 square lattice for a random initial condition with equal densities of spin up and spin down. Figure courtesy of V. Spirin.

(a)   Ising–Glauber   (b)   voter

Fig. 8.7.   Comparison of the rates of an update event in (a) the Ising–Glauber model at zero temperature and (b) the voter model on the triangular lattice.

factorized as a product of soluble one-dimensional equations. In contrast, for $d > 1$ the spin variables inside the hyperbolic tangent function in Eq. (8.77) lead to an infinite equation hierarchy for correlation functions that cannot be closed. The case $d = 1$ is exceptional because we can write the flip rate as the quadratic form in Eq. (8.43), which ultimately leads to solvable rate equations for the mean spin and the spin correlation function.

Example 8.3. *Flip rates of the Ising–Glauber model at zero temperature*. Invoking locality, left/right, and up/down symmetries as in one dimension (Section 8.5), the most general local flip rate of a spin s_i must have the form

$$w_i = a + b\,s_i \sum_{\alpha \in \langle i \rangle} s_\alpha + c\,s_i \sum_{\alpha,\beta,\gamma \in \langle i \rangle} s_\alpha s_\beta s_\gamma.$$

The first sum is over the four nearest neighbors of $i = (x, y)$, i.e. $(x \pm 1, y)$ and $(x, y \pm 1)$; the second sum is over the four possible ways of choosing three mutually distinct nearest neighbors of i, i.e. $\alpha \neq \beta, \beta \neq \gamma, \gamma \neq \alpha$. We now impose the following conditions on the flip rate for zero-temperature dynamics (Fig. 8.8):

$$a - 4b - 4c = 1, \qquad \text{Fig. 8.8(a)},$$
$$a - 2b + 2c = 1, \qquad \text{Fig. 8.8(b)},$$
$$a = 1/2, \qquad \text{Fig. 8.8(c)}.$$

(There are another three degenerate equations that arise from similar configurations to those in Fig. 8.8, but with the central spin s_i reversed.) These constraints give the flip rate

$$w_i = \frac{1}{2} \left[1 - \frac{3}{8} s_i \sum_{\alpha \in \langle i \rangle} s_\alpha + \frac{1}{8} s_i \sum_{\alpha,\beta,\gamma \in \langle i \rangle} s_\alpha s_\beta s_\gamma \right]. \tag{8.78}$$

With this flip rate we can determine the evolution of the correlation functions. For example, combining (8.49a) and (8.78) gives

$$\frac{dS_i}{dt} = -S_i + \frac{3}{8} \sum_{\alpha \in \langle i \rangle} S_\alpha - \frac{1}{8} \sum_{\alpha,\beta,\gamma \in \langle i \rangle} S_{\alpha\beta\gamma}, \tag{8.79}$$

which couples the one-spin correlation function $S_i = \langle s_i \rangle$ to the three-spin correlation function $S_{\alpha\beta\gamma} = \langle s_\alpha s_\beta s_\gamma \rangle$. Therefore we need equations for the three-spin correlation functions $S_{\alpha\beta\gamma}$. These equations contain five-spin correlation functions, leading to a never-ending spiral. In general the evolution equations for the correlation functions are hierarchical, rather
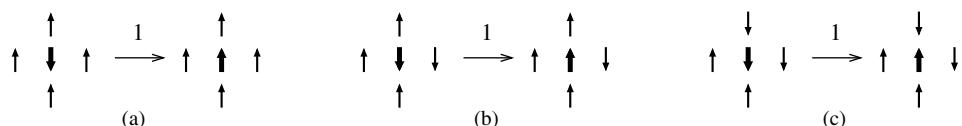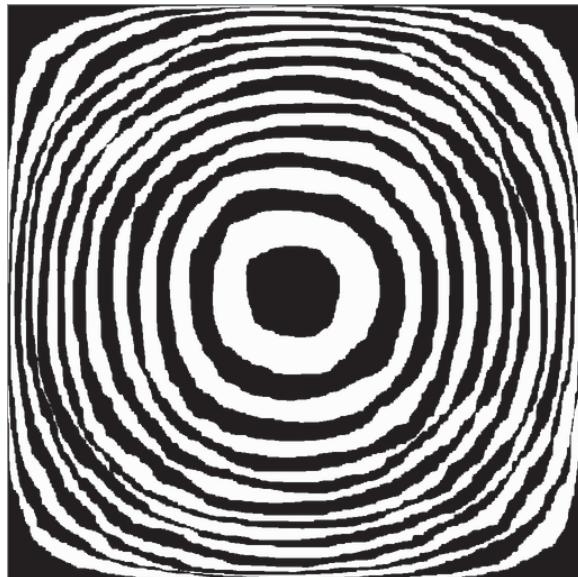
Distinct spin configurations on the square lattice that are used to determine the flip rates.

Fig. 8.9.   Evolution of an initially square Ising droplet. The figure is from R. Cerf and S. Louhichi, *Prob. Theory Rel. Fields* **137**, 379 (2007). Reprinted with permission from R. Cerf.

than recursive, in nature. Such hierarchical equations are generally intractable and therefore Glauber's spin-flip dynamics in two dimensions is unsolvable.

A key feature of zero-temperature Glauber dynamics in greater than one dimension is that the average magnetization is *not* conserved. Mathematically, non-conservation arises because the sum in Eq. (8.79) over all sites (and its generalization to higher dimensions) does not vanish. An intuitive understanding of this non-conservation is provided by the evolution of an isolated droplet of radius R that consists of one phase in a sea of the opposite phase (Fig. 8.9). The majority rule nature of the Glauber dynamics leads to an effective surface tension so that the radius $R(t)$ of an isolated droplet shrinks as $R(t)^2 = R_0^2 - at$ (a continuum description of this shrinking will be given in Section 9.3). For the voter model, in contrast, magnetization conservation causes an isolated droplet to become more diffuse and there is no systematic shrinking (Fig. 8.1).

The long-time properties of the Ising–Glauber model are particularly intriguing. When an initially finite random system of linear dimension L is suddenly quenched to zero temperature, the final state depends fundamentally on the spatial dimensionality. In one dimension, a finite system ultimately reaches complete alignment – the ground state. In two dimensions, the ground state is reached with probability p that is strictly less than 1, while states that consist of an even number of perfectly straight stripes of width ≥ 2 (Fig. 8.10) are reached with probability $1 - p$.[14] These stripe states are static and represent a possible final

---

[14] We always tacitly assume that the initial magnetization $m_0 = 0$. If $m_0 \neq 0$, the system always reaches the field-aligned ground state in the thermodynamic limit for any spatial dimension.
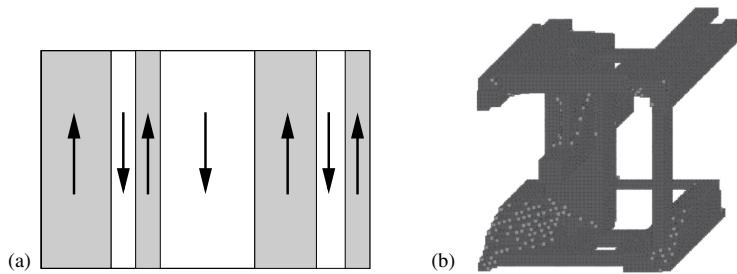
(a) A metastable configuration on a periodic rectangular lattice that consists of six vertical stripes. (b) A metastable configuration on the simple cubic lattice; this figure courtesy of V. Spirin.
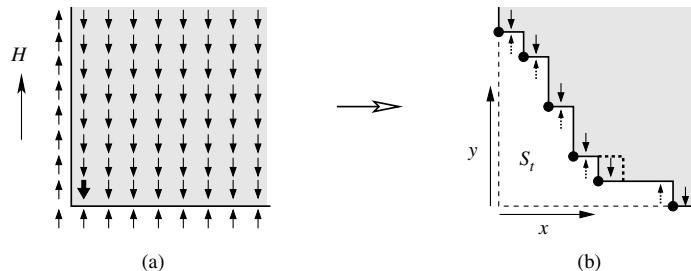
(a) A corner interface in (a) the initial state and (b) at $t > 0$. The unique flippable spin in the initial state is indicated by the thick arrow in (a). The evolving interface encloses an area S_t at time t. In (b), there is always one more flippable ↓ spin (solid) and flippable ↑ spin (dashed) in zero magnetic field.

state of the evolution. The total number $\mathcal{M}$ of these metastable states is huge: $\ln \mathcal{M} \sim \sqrt{N}$ (problem 8.15).

In greater than two dimensions, the ground state is never reached,[15] and the full nature of the metastable states is not known. Typical metastable states can be viewed as the discrete analogs of *minimal surfaces* (Fig. 8.10, and see Fig. 9.4 for continuum analogs), in which the interface between the two phases consists of both positively curved and negatively curved portions such that the entire interface is static. Strangely, the final state is not an isolated metastable state, but rather a connected iso-energy set within which the system wanders forever by zero-temperature Glauber spin-flip dynamics.

Example 8.4. *The corner interface*. The corner interface on the square lattice (Fig. 8.11) is perhaps the simplest example of a two-dimensional interface whose evolution by Glauber dynamics can be solved exactly. The corner is initially at the origin, so that the spins in the first quadrant point down and all other spins point up. Initially only the corner spin at $(0, 0)$ can flip because it is the only spin in a zero-energy environment. After this flip, the two neighboring spins at $(1, 0)$ and $(0, 1)$ can flip, while the corner spin at $(0, 0)$ can flip back to its initial orientation. This pattern continues – at any time, a certain number

---

[15] By never, we mean that the probability of reaching the ground state decays extremely rapidly with L.
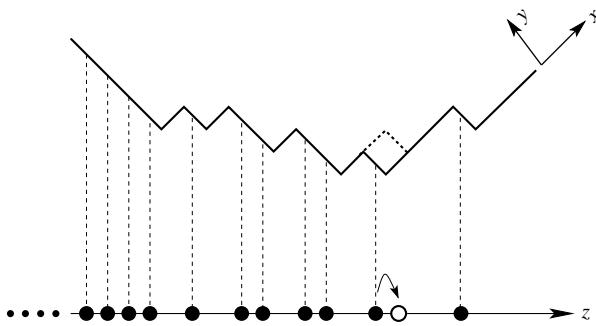
Fig. 8.12.   Correspondence between the corner interface, rotated by $\pi/4$, and the ASEP. A spin-flip event is shown and its correspondence in the ASEP.

N of spins can flip into the majority phase and exactly $N - 1$ spins can flip back into the minority phase. All these spin flips are energy-conserving, so that the interface length remains constant. Moreover the interface keeps its integrity; the creation of droplets or the calving of droplets from the initial interface is impossible. These features make the corner interface problem tractable.

There are two types of dynamics that depend on the magnetic field:

Driven interface (weak positive field $0 < H < 2J$). Here $\downarrow \to \uparrow$ events occur only at corners, while $\uparrow \to \downarrow$ events never occur. In Fig. 8.11(b), for example, there are six locations where $\downarrow \to \uparrow$ spin flips can happen. To find the interface shape, we first map the spin dynamics to the asymmetric exclusion process (ASEP) discussed in Chapter 4. The mapping is performed in two steps. First we rotate the corner about the origin counterclockwise by $\pi/4$. Next, we project a northwest/southeast segment to an occupied particle on the horizontal axis, while a northeast/southwest segment maps to an unoccupied site (Fig. 8.12). A spin-flip event corresponds to changing a section of the interface from $\llcorner$ to $\urcorner$. In the equivalent ASEP, this update corresponds to a particle hopping to the right, $\bullet \circ \to \circ \bullet$, which can occur only if the right neighbor is vacant. Because each eligible particle hops at the same rate, the interface evolution coincides with the ASEP.

By virtue of the $\pi/4$ tilt between the interface and the equivalent ASEP, a spin-flip event at (x, y) on the interface maps to the point $z = x - y$ on the tilted line. The presence of a particle at z on the line corresponds to the interface height at (x, y) decreasing by 1. In the continuum limit, the interface slope and particle density in the corresponding ASEP are therefore related by $y'(x) = -n(z)$. The initial corner interface in the spin system is equivalent, in the corresponding ASEP, to the half-line $z < 0$ being fully occupied, $n(z, t = 0) = 1$ for $z < 0$, and the half-line $z > 0$ being vacant, $n(z, t = 0) = 0$ for $z > 0$. By calculating the smoothing of this density step in the ASEP, we can thereby infer the shape of the equivalent interface.

In the continuum limit, the particle density $n(z, t)$ on the line in the ASEP obeys the Burgers-like equation

$$\frac{\partial n}{\partial t} + \frac{\partial [n(1 - n)]}{\partial z} = \frac{\partial^2 n}{\partial z^2} \,, \tag{8.80}$$

that we encountered previously in Chapter 4. We may ignore the diffusive term in the long-time limit because its role is subdominant compared to the influence of the constant rightward bias. The solution to Eq. (8.80) without the diffusive term was given in Example 4.3 in Section 4.3, from which the particle density is

$$n(z,t) = \begin{cases} 1, & z < -t, \\ \frac{1}{2}\left(1 - z/t\right), & |z| < t, \\ 0, & z > t. \end{cases} \tag{8.81}$$

From $n(z) = -y'(x)$, the locus for the interface is then given by

$$y(x,t) = \int_{x-y}^{\infty} n(z,t)\, dz = \int_{\max(x-y,-t)}^{t} n(z,t)\, dz. \tag{8.82}$$

Performing this integral then gives the implicit equation

$$y = \begin{cases} \frac{1}{2}\left[\frac{t}{2} - (x-y) + \frac{1}{2t}(x-y)^2\right], & |x-y| < t, \\ 0, & x-y > t, \end{cases} \tag{8.83}$$

that reduces, for $0 < x, y < t$, to the remarkably simple parabolic form for the interface shape:

$$\sqrt{x} + \sqrt{y} = \sqrt{t} \tag{8.84}$$

for $0 < x, y < t$. Outside of this range the interface is unperturbed.

Ising interface (zero field). The spin-flip events $\downarrow \to \uparrow$ and $\uparrow \to \downarrow$ can occur with the same rate for all corner spins. Because there is always one more flippable $\downarrow$ spin than $\uparrow$ spin (Fig. 8.11(b)), there is a slight constant bias in the interface motion. Following the same reasoning as for the driven interface, the interface evolution is now equivalent to the symmetric exclusion process on the tilted line. However, as discussed in Chapter 4, exclusion does not affect the overall density. Consequently, the particle density satisfies the standard diffusion equation

$$\frac{\partial n}{\partial t} = \frac{\partial^2 n}{\partial z^2}$$

whose solution for the step initial condition is

$$n(z,t) = \frac{1}{\sqrt{\pi}} \int_{z/\sqrt{4t}}^{\infty} e^{-w^2}\, dw = \frac{1}{2}\,\text{erfc}\left(\frac{z}{\sqrt{4t}}\right), \tag{8.85}$$

where $\text{erfc}(x)$ is the complementary error function. While the particle profile is easy to obtain, the transformation back to the interface shape via Eq. (8.82) does not result in an explicit and simple result such as (8.84). Nevertheless, the interface is formally solved and several of its features can be expressed explicitly; e.g. the intersection of the interface with the diagonal $y = x$ occurs at $x = y = \sqrt{t/\pi}$.

8.7 Spin-exchange dynamics

As mentioned at the outset of Section 8.3, there are two fundamental classes of spin dynamics: magnetization conserving and magnetization non-conserving. The former class is appropriate to describe alloy systems, where the two different spin states correspond to the two component atoms that comprise the alloy. In the phase separation of an alloy into domains of pure metal, a plausible dynamics is that the positions of atoms of different species are exchanged. In the spin language this evolution corresponds to the exchange of antiparallel neighboring spins:

$$\cdots \underline{\uparrow\downarrow} \cdots \longrightarrow \cdots \underline{\downarrow\uparrow} \cdots .$$

This evolution, known as Kawasaki spin-exchange dynamics, *strictly* conserves the magnetization in every update event. This strict conservation has far-reaching consequences that will become clearer when we discuss continuum theories of spin dynamics in the next chapter.

The transition rate

There are three types of update events: energy-raising, energy-lowering, and energy-neutral. As illustrated in Fig. 8.13, the energy-neutral update arises by the simultaneous hopping of two nearest-neighbor domain walls; equivalently, this update occurs when an isolated down spin in a sea of up spins exchanges its position with a neighboring spin. As long as the domain-wall pair remains isolated from all other domain walls, the pair hops freely between neighboring lattice sites. Such pairs can be viewed as the elementary excitations of the spin system. The hopping rate of a domain-wall pair merely sets the time scale, so there is no loss of generality in setting this rate to $1/2$, as in Glauber dynamics. Because such diffusive moves do not alter the energy, they automatically satisfy the detailed balance condition. The rates of the remaining two update events are then set by detailed balance. Since spin exchange involves the interactions among four spins – the two that flip and their two neighbors – the rates depend on the total energy of the three bonds connecting these

(a)   (b)   (c)   (d)
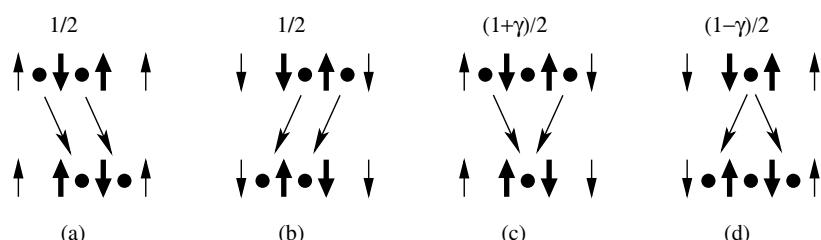
Fig. 8.13. Energy-neutral update events (a) and (b), energy-lowering events (c), and energy-raising events (d) for Kawasaki dynamics in one dimension. The spins that flip are shown bold. Also shown are the corresponding domain-wall particles (•) and the transition rates for these four events.

four spins. The detailed balance condition is then

$$\frac{w_3}{w_{-1}} = \frac{p_{-1}}{p_3} = \exp(4\beta J), \tag{8.86}$$

where w_q is the transition rate out of a state with energy qJ and p_q is the equilibrium probability of this state. Using the Glauber notations in which $w_3 = (1 + \gamma)/2$ and $w_{-1} = (1 - \gamma)/2$ for energy-raising and energy-lowering transitions respectively, the detailed balance condition has the same form as in Glauber dynamics, $(1 + \gamma)/(1 - \gamma) = \exp(4\beta J)$, or $\gamma = \tanh 2\beta J$.

To determine the transition rates, we first must guarantee that spins i and $i + 1$ are antiparallel. This constraint is achieved by the factor $(1 - s_i s_{i+1})/2$ that equals $+1$ if the two spins are antiparallel and equals zero otherwise. The form of the flip rate then depends on the interaction energy between the pairs s_{i-1} and s_i, and between s_{i+1} and s_{i+2}. The flip rate should be a symmetric function of these two bond energies and the rate should be proportional to $(1 + \gamma)/2$, $1/2$, and $(1 - \gamma)/2$ respectively, when the signs of these bond energies are $--$, $+-$, and $++$. These constraints lead to the transition rate

$$w_i(s_i, s_{i+1}) = \frac{1}{2}\left[1 - \frac{\gamma}{2}\left(s_{i-1}s_i + s_{i+1}s_{i+2}\right)\right] \times \frac{1}{2}\left(1 - s_i s_{i+1}\right). \tag{8.87}$$

With this rate, the evolution of spin correlation functions is not closed. One-spin averages are coupled to three-spin averages, two-spin averages are coupled to four-spin averages, etc. Thus we again find the technical roadblock that the equation of motion for a particular correlation function generates an infinite hierarchy of equations for high-order correlations. It requires considerable technical effort and artistry to find a tractable and accurate scheme to truncate such infinite hierarchies into a finite set of soluble equations.

Frustration at zero temperature

Because of the magnetization conservation in Kawasaki dynamics, a system that is quenched to zero temperature must always get stuck in a metastable state. The metastable states are characterized by each domain-wall particle being separated by more than a nearest-neighbor distance from any other domain wall. Equivalently the lengths of all spin domains are two or longer. The number of such configurations in a system of length L asymptotically grows as g^L, where $g = (1 + \sqrt{5})/2$ is the golden ratio. At zero temperature these metastable states prevent the system from reaching the ground state. At non-zero temperature, these states merely slow the approach toward equilibrium.

Let's now study this frustration at zero temperature (only energy-lowering transitions allowed) for the simple case of the antiferromagnetic initial state (Fig. 8.14). There are slight differences when energy-conserving transitions (diffusion of neighboring walls) can also occur, but the former is simpler to treat analytically. It is helpful to visualize the evolution in terms of the domain walls that occupy the sites of the dual lattice. At zero temperature, the Kawasaki spin-exchange update step is equivalent to picking three contiguous domain-wall particles at random and removing the two outside particles. This converts any triplet ●●● to
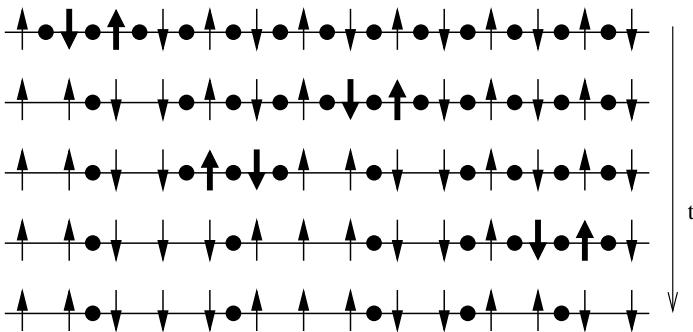
Fig. 8.14. Evolution of the antiferromagnetic initial state by energy-lowering Kawasaki dynamics. The spins that are about to flip are highlighted. Also shown are the domain walls between neighboring anti-aligned spins.

○ ● ○ equiprobably. We now use the tools of random sequential adsorption (see Section 7.1) to solve the problem.

Let E_k be the probability that an interval of k consecutive sites (in the dual lattice) are all occupied by domain walls. This probability evolves by

$$\frac{dE_k}{dt} = -(k-2)E_k - 2E_{k+1} - 2E_{k+2} \qquad (8.88)$$

for $k \geq 3$. This equation reflects the different ways that the transition $● ● ● \rightarrow ○ ● ○$ can occur and alter the number of occupied intervals of length k. There are $k-2$ ways for this transition to occur in the interior of a k-interval. There are also two ways for this transition to occur with two sites at the edge of the k-interval and one site outside; in this case, the event must occur in an occupied interval of length $k+1$. Finally, there are also two ways with one site at the edge of the k-interval and two sites outside the k-interval.

Having formulated the problem in terms of the evolution of occupied intervals, we now apply the tools developed in Section 7.1 to solve Eq. (8.88). We make the exponential ansatz $E_k = \Phi(t)\,e^{-(k-2)t}$ to reduce the infinite set of equations in (8.88) to the single equation (see the discussion accompanying Eq. (7.4))

$$\frac{d\phi}{dt} = -2\phi\left(e^{-t} + e^{-2t}\right). \qquad (8.89)$$

For the antiferromagnetic initial condition, the dual lattice is completely occupied. Thus $E_k = 1$ initially, so that $\Phi(0) = 1$. Integrating (8.89) subject to this initial condition gives the interval probabilities for $k \geq 2$,

$$E_k(t) = \exp\left[-(k-2)t + e^{-2t} + 2e^{-t} - 3\right]. \qquad (8.90)$$

Since two domain-wall particles are lost in each update event and these events occur with rate E_3, the domain-wall density $\rho \equiv E_1$ satisfies $d\rho/dt = -2E_3$. Using Eq. (8.90) for E_3

and integrating, the domain-wall density is

$$\rho(t) = 1 - 2 \int_0^t du \, \exp\left[-u + e^{-2u} + 2e^{-u} - 3\right], \qquad (8.91)$$

from which the final jamming density is $\rho_{jam} \equiv \rho(\infty) \doteq 0.450898$. Thus there is incomplete relaxation, as the density of domain walls is close to $\frac{1}{2}$ in the final jammed state. Moreover, the relaxation to the jamming density is exponential in time,

$$\rho(t) - \rho_{jam} \simeq e^{-3} \, e^{-t} . \qquad (8.92)$$

Coarsening at infinitesimal temperature

While the one-dimensional chain with Kawasaki dynamics quickly reaches a jammed state at zero temperature, the equilibrium state *is* reached for any non-zero temperature, no matter how small. Because the equilibrium correlation length $\xi = [\ln(\coth \beta J)]^{-1} \sim e^{2\beta J}$ (see Eq. (8.55)) diverges as the temperature approaches zero, one can set the temperature sufficiently small so that the correlation length is much larger than the length of the system. Then the equilibrium state consists of a single domain of ferromagnetic order so that the evolution resembles zero-temperature coarsening.

The large separation of time scales between energy-raising updates and all other updates leads to an appealing description of the domain evolution within an extremal dynamics framework. Since the rate of an energy-raising update equals $e^{-4\beta J}$, the typical time for such an event is $\tau \equiv e^{4\beta J}$. We now define the elemental time unit to be τ. In this time unit, energy-neutral and energy-lowering events occur instantaneously. Starting from an initial state, the system instantly reaches a frustrated state in which further energy-neutral or energy-lowering moves are not possible. After a time τ has elapsed (on average) an energy-raising event occurs that is subsequently followed by a burst of energy-neutral and energy-lowering events until the system reaches another frustrated metastable state. This pattern of an energy-raising event followed by a burst of complementary events continues until a finite system reaches a single-domain state. As we now show, this dynamics leads to the typical domain size growing in time as $t^{1/3}$, a general feature of order-parameter-conserving dynamics. One of the appealing features of Kawasaki dynamics in one dimension is that this $t^{1/3}$ coarsening emerges in a direct way. In contrast, we will see in the next chapter that it requires much more work to deduce the $t^{1/3}$ coarsening from continuum approaches.

At some intermediate but still long time, the system reaches a low-energy state that consists of alternating domains of typical length ℓ. The subsequent evolution at low temperature is controlled by the rare, energy-raising updates, that require a time $\tau = e^{4\beta J}$, where a pair of domain walls nucleates around an existing isolated domain wall. Once this domain-wall triplet forms, two neighboring walls in the triplet can diffuse freely as a bound domain-wall pair until another isolated domain wall is encountered. This domain wall could be the original wall that was left behind when the doublet began diffusing, or this wall could be at the next interface. In the latter case, we can view the spin as moving from one interface to the next so that one domain increases its size by one and another shrinks by one (Fig. 8.15). The
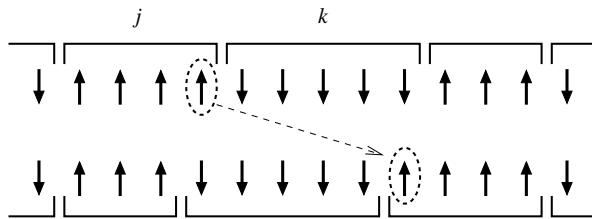
Fig. 8.15. Effective domain diffusion by Kawasaki dynamics at infinitesimal temperature. A ↓ spin from the j-domain (dashed oval) splits off and eventually reaches the right edge of the neighboring k-domain with probability $1/k$. The net result of the diffusion of the spin across the k-domain is that the k-domain moves rigidly to the left by one lattice spacing.
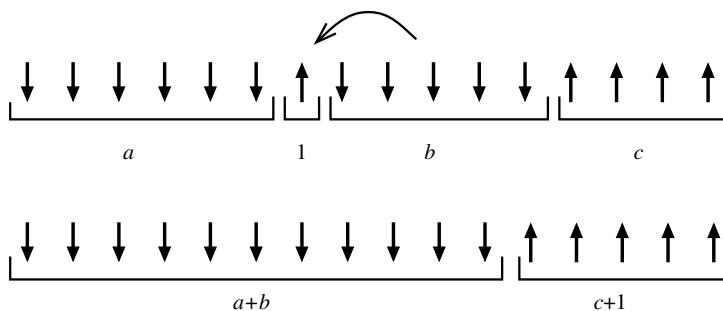
Fig. 8.16.   The outcome after domain merging.

probability that the spin moves to an adjacent interface is a first-passage problem that was discussed in Section 2.5. Once the spin has split off, it is a distance 1 from its initial interface and a distance $k - 1$ from the interface on the other side. Since the spin diffuses nearly freely,[16] it eventually reaches the other side with probability $1/(2k - 2)$ and returns to its starting position with probability $1 - 1/(2k - 2)$. Hence the probability that the k-domain hops by one lattice spacing scales as k^{-1}, corresponding to a domain diffusion coefficient $D(k) \sim k^{-1}$.

Thus in the low-temperature limit, the spin dynamics maps to an effective isotropic hopping of entire domains by one step to the left or the right[17] (Fig. 8.15). However, domains of length 1 disappear whenever one of their neighboring domains hops toward them, so that four neighboring domains merge into two (Fig. 8.16). The net effect of this last process is domain coarsening. We now estimate the rate of this coarsening by a scaling argument. Since each domain performs a random walk, coalescence occurs whenever a domain diffuses of the order of its own length. In such an event, a domain of typical length ℓ grows by an amount $\Delta \ell$ that is also of the order of ℓ, while the time between coalescence

---

[16] The hopping of the spin is not symmetric when it is adjacent to either its starting position or the other side of the domain (see problem 8.17).

[17] There is an anomaly associated with domains of length 2 that can be ignored for the purposes of this discussion.

events is $\Delta t \sim \tau \ell^2/D(\ell)$. Thus

$$\frac{\Delta \ell}{\Delta t} \sim \frac{\ell}{\tau \ell^2/D(\ell)} \sim \frac{1}{\tau \ell^2} \,,$$

so that domains grow as

$$\ell \sim (t/\tau)^{1/3} \,. \tag{8.93}$$

Since the system is maintained at small but positive temperature, equilibrium is eventually reached. The crossover time between coarsening and achieving equilibrium can be estimated by equating the typical domain size and the equilibrium correlation length, $(t_*/\tau)^{1/3} \sim \xi$. This gives the extraordinarily long time $t_* \sim \tau \xi^3 \sim e^{10\beta J}$ at low temperature.

It is customary to define the *dynamical exponent z* in terms of the growth of the typical length scale in coarsening via $\ell \sim t^z$. For the non-conserved Glauber and the conserved Kawasaki dynamics, the dynamical exponent is

$$z = \begin{cases} 1/2, & \text{non-conservative dynamics,} \\ 1/3, & \text{conservative dynamics.} \end{cases} \tag{8.94}$$

While we have derived these results in one dimension, they are generic for all spatial dimensions. Conservation laws strongly constrain the nature of non-equilibrium dynamics.

8.8 Cluster dynamics

Thus far, we have discussed the examples of *local* Glauber single-spin flip dynamics and the Kawasaki spin-exchange dynamics. In spite of their idealized natures, these rules have formed the basis for many simulational studies of coarsening and dynamic critical phenomena because of their apparent connection to the evolution of real systems. However, dynamics that is based on flipping single spins or two neighboring spins is computationally inefficient. Compounding this inefficiency, the dynamics significantly slows down close to criticality. To mitigate these drawbacks, dynamical update rules have been proposed in which an entire suitably defined cluster of spins simultaneously flips, but at the same time still preserves the correct equilibrium properties. Because of their efficiency, such *cluster algorithms* have been used extensively to simulate the equilibrium behavior of many-body statistical mechanical and lattice field theory models. The Swendsen–Wang and the Wolff algorithms are two of the earliest and most prominent examples of cluster dynamics.

As we now discuss, the master equations that encode these algorithms are analytically soluble in one dimension. These two examples provide nice illustrations of the power of the master equation approach.

Swendsen–Wang algorithm

In the Swendsen–Wang algorithm, an entire domain of aligned spins is chosen at random and all its spins flip simultaneously, as illustrated by

$$\cdots \uparrow\uparrow\underbrace{\downarrow\downarrow\downarrow\downarrow}\uparrow\uparrow \cdots \longrightarrow \cdots \uparrow\uparrow\underbrace{\uparrow\uparrow\uparrow\uparrow}\uparrow\uparrow \cdots .$$

In each update, the energy always decreases and there is a concomitant loss of two domains, so that their number density ρ decreases according to $d\rho/dt = -2\rho$; here we take the flip rate to be 1 without loss of generality. For the antiferromagnetic initial condition, $\rho(0) = 1$, so that the domain density $\rho(t) = e^{-2t}$. Since the average domain length $\langle \ell \rangle$ is the inverse of the domain density, we have $\langle \ell \rangle = e^{2t}$. When this average length reaches the system length L the dynamics is complete. From this criterion, we obtain the time to reach the ground state as $T_L \sim \ln L$.

Let's now consider the domain length distribution. We define c_ℓ as the density of domains of length ℓ. When a domain is flipped, it merges with its two neighbors, so that the length of the resulting domain equals the length of these three initial domains. As a result of this three-body aggregation, c_ℓ evolves according to

$$\frac{dc_\ell}{dt} = -3c_\ell + \frac{1}{\rho^2} \sum_{i+j+k=\ell} c_i\, c_j\, c_k . \tag{8.95}$$

The factor of $-3c_\ell$ accounts for domain loss when a domain of length ℓ or either of its neighboring domains flips. The last term accounts for the gain in c_ℓ due to the flipping of a domain of length j that then merges with its two neighboring domains of lengths i and k, with $\ell = i + j + k$. The probability for this merging event equals $c_i c_k/\rho^2$, which is the probability that the j-domain lies between domains of lengths i and k. Notice that newly created domains do not affect their neighbors, nor are they affected by their neighbors. Thus if the domains are initially uncorrelated, they remain uncorrelated. Because no spatial correlations are generated, Eq. (8.95) is exact.

We can obtain a cleaner-looking equation by introducing $P_\ell \equiv c_\ell/\rho$, the probability that a domain of length ℓ exists. By construction, $\sum_\ell P_\ell = 1$. By expressing Eq. (8.95) in terms of P_ℓ and using the equation of motion for the density $\dot{\rho} = -2\rho$, we find that the evolution equation for P_ℓ is

$$\frac{dP_\ell}{dt} = -P_\ell + \sum_{i+j+k=\ell} P_i P_j P_k . \tag{8.96}$$

As in many previous examples, it is expedient to introduce the generating function $\mathcal{P}(z) = \sum_\ell P_\ell z^\ell$ to simplify the convolution form of the gain term and thereby reduce (8.96) to $\partial \mathcal{P}/\partial t = -\mathcal{P} + \mathcal{P}^3$. Writing $1/(\mathcal{P}^3 - \mathcal{P})$ in a partial fraction expansion, the equation for $\partial \mathcal{P}/\partial t$ can be integrated by elementary methods and the solution is

$$\mathcal{P}(z, t) = \frac{\mathcal{P}_0(z)e^{-t}}{\sqrt{1 - \mathcal{P}_0(z)^2(1 - e^{-2t})}} , \tag{8.97}$$

where $\mathcal{P}_0(z)$ is the initial generating function.

For the simple case of the antiferromagnetic initial condition, all initial domains have length 1 and the corresponding initial condition for the generating function is $\mathcal{P}_0(z) = z$. Now we expand the generating function in (8.97) in powers of z to yield the domain number distribution

$$P_{2\ell+1} = \binom{2\ell}{\ell} \left(\frac{1 - e^{-2t}}{4} \right)^{\ell} e^{-t} . \tag{8.98}$$

Notice that this generating function leads to domains that have odd lengths only. From this distribution, the average domain length grows exponentially with time, $\langle \ell \rangle = e^{2t}$, as found previously. Finally, using Stirling's approximation, the asymptotic length distribution approaches the scaling form $P_\ell \to e^{-2t} \, \Phi(\ell e^{-2t})$ with the scaling function

$$\Phi(x) = \frac{1}{\sqrt{2\pi x}} e^{-x/2} . \tag{8.99}$$

In scaled units, the distribution is dominated by the singularity in the number of small domains.

Wolff algorithm

In the Wolff cluster algorithm, a spin is selected at random and the domain that it belongs to is flipped. This protocol further accelerates the dynamics compared to the Swendsen–Wang algorithm because larger domains are more likely to be updated. Schematically, the Wolff dynamics may be represented as

$$\cdots \uparrow\uparrow \underbrace{\downarrow\downarrow \cdots \downarrow\downarrow}_{k} \uparrow\uparrow \cdots \xrightarrow{k} \cdots \uparrow \underbrace{\uparrow\uparrow \cdots \uparrow\uparrow}_{k} \uparrow\uparrow \cdots ,$$

so that a flipped domain again merges with its neighbors. Since each spin is selected randomly, the time increment associated with any update is identical. The domain density therefore decreases with constant rate $\dot{\rho} = -2$. Consequently, $\rho(t) = 1 - 2t$ so that the entire system is transformed into a single domain in a finite time $t_c = 1/2$. Correspondingly, the average domain length, $\langle \ell \rangle = (1 - 2t)^{-1}$, diverges as $t \to t_c$.

The evolution of the domain length distribution is governed by the direct generalization of (8.96):

$$\frac{dP_\ell}{dt} = -\ell P_\ell + \sum_{i+j+k=\ell} j P_i P_j P_k . \tag{8.100}$$

The generating function $\mathcal{P}(z,t) = \sum_\ell P_\ell z^\ell$ now satisfies

$$\frac{\partial \mathcal{P}}{\partial t} = z \, (\mathcal{P}^2 - 1) \, \frac{\partial \mathcal{P}}{\partial z} . \tag{8.101}$$

To solve this equation, we first transform from the variables (t, z) to $(\tau, y) \equiv (t, t - \ln z)$ to write the linear terms as a single derivative and give

$$\frac{\partial \mathcal{P}}{\partial \tau} = -\mathcal{P}^2 \, \frac{\partial \mathcal{P}}{\partial y} . \tag{8.102}$$

We now employ the same procedure as that used in the solution of aggregation with the product kernel (see the discussion leading to Eq. (5.41) in Chapter 5). Along the characteristic curve defined by $\partial y/\partial \tau = \mathcal{P}^2$, $\mathcal{P}$ is constant. Thus $y = \mathcal{P}^2 \tau + G(\mathcal{P})$, or equivalently

$$t - \ln z = \mathcal{P}^2 t + G(\mathcal{P}), \tag{8.103}$$

with $G(\mathcal{P})$ determined by the initial condition. For the antiferromagnetic initial state, $\mathcal{P}_0(z) = z$, so that $G(\mathcal{P}) = -\ln \mathcal{P}$. Substituting $G(\mathcal{P}) = -\ln \mathcal{P}$ into (8.103) and exponentiating yields the following implicit equation for the generating function:

$$z = \mathcal{P} \, e^{t - \mathcal{P}^2 t}. \tag{8.104}$$

The length distribution P_ℓ is just the ℓth term in the power series expansion of $\mathcal{P}(z)$. Formally, this term may be extracted by writing P_ℓ as the contour integral

$$P_\ell = \frac{1}{2\pi i} \oint \frac{\mathcal{P}(z)}{z^{\ell+1}} \, dz,$$

then transforming the integration variable from z to $\mathcal{P}$, and using the Lagrange inversion formula (see the highlight on page 148 in Chapter 5). These steps give

$$P_\ell = \frac{1}{2\pi i} \oint \frac{\mathcal{P}(z)}{z^{\ell+1}} \, dz = \frac{1}{2\pi i} \oint \frac{\mathcal{P}}{z(\mathcal{P})^{\ell+1}} \frac{dz}{d\mathcal{P}} \, d\mathcal{P},$$

$$= \frac{e^{-\ell t}}{2\pi i} \oint e^{\ell \mathcal{P}^2 t} \left[\frac{1}{\mathcal{P}^\ell} - \frac{2t}{\mathcal{P}^{\ell-2}} \right] d\mathcal{P}, \tag{8.105}$$

where we use $dz/d\mathcal{P} = e^{t-\mathcal{P}^2 t}(1-2\mathcal{P}^2 t)$ in the above integral. Now we find the residues by expanding $e^{\ell \mathcal{P}^2 t}$ in a power series and keeping only the coefficient of $1/\mathcal{P}$ in the integrand. Because the power series is even in $\mathcal{P}$, only P_ℓ for odd values of ℓ is non-zero and we find

$$P_\ell = e^{-\ell t} \left[\frac{(\ell t)^{(\ell-1)/2}}{[(\ell-1)/2]!} - 2t \frac{(\ell t)^{(\ell-3)/2}}{[(\ell-3)/2]!} \right].$$

After some algebra, the domain length distribution is

$$P_{2\ell+1}(t) = \frac{(2\ell+1)^{\ell-1}}{\ell!} \, t^\ell \, e^{-(2\ell+1)t}. \tag{8.106}$$

8.9 Notes

Despite the central role that kinetic spin systems have played in non-equilibrium statistical physics, there are no books for physicists on Glauber dynamics or related microscopic models of spin evolution. Two useful reviews about the basic phenomenology of non-equilibrium spin systems, mostly from the continuum perspective, are given in [139, 140].

The study of the voter model started in the probability theory community and a basic reference is the monograph by Liggett [141]. The simulation study that illustrated the absence of a surface tension in the voter model, as illustrated in Fig. 8.1, appeared in [142]. The exact solution of the voter model using techniques from non-equilibrium statistical physics is presented in [143, 144]. An excellent starting point for learning about the Glauber dynamics is the beautiful article by Glauber himself [145]. References [146, 147] make the next step after Glauber and analyze higher-order correlation functions, a subject that we touched upon in Example 8.1. The temporal evolution of the domain length distribution for the zero-temperature Ising–Glauber chain is still unknown in full detail, but this challenging problem ultimately seems tractable. However, several basic features of this distribution have been established using the independent interval approximation [148] and by exact approaches [149]. The independent interval approximation is useful for dealing with many-body problems in one dimension and sometimes it is even asymptotically exact [150].

The zero-temperature evolution of the Ising–Glauber model and the phenomenon of freezing into metastable states was studied in [151, 152]. The exact evolution of single interfaces by Glauber kinetics, such as the corner problem, was solved by Rost [153], using similar techniques to those employed here. More about the shrinking of a square droplet can be found in [154–156]. Spin-exchange dynamics that explicitly conserves the magnetization was introduced by Kawasaki [157]. Reference [158] developed the domain picture for the $t^{1/3}$ coarsening of domains with low-temperature Kawasaki dynamics. The idea for drastically accelerating simulations of kinetic Ising models by flipping correlated groups of spins was formulated by Swendsen and Wang [159], while the related Wolff dynamics was proposed in [160]; the general topic of acceleration algorithms for kinetic Ising simulations is reviewed in [161]. The Swendsen–Wang and Wolff dynamics for the Ising chain were discussed in [162, 163].

8.10 Problems

8.1 Consider the voter model in one dimension.
   (a) Suppose that the initial state consists of two $+1$ voters in a sea of "uncommitted" voters, namely, voters that may be in either opinion state with the same probability: $S(x, t = 0) = \delta_{x,-1} + \delta_{x,1}$.
      (i) Determine $S(x, t)$ for all integer x.
      (ii) Show that $S(0, t) = 2e^{-t}I_1(t)$. Using the definition of the Bessel function $I_1(t)$ prove the validity of the (physically obvious) inequality $S(0, t) < 1$.
   (b) Consider now the "dipole" initial condition, $S(x, t = 0) = \delta_{x,1} - \delta_{x,-1}$, namely one $+1$ voter and one -1 voter in a sea of uncommitted voters. For this initial condition show that

$$S(x, t) = \frac{2x}{t} e^{-t}I_x(t).$$

8.2 This problem concerns the derivation of the exact solution for the pair correlation function in the voter model in arbitrary spatial dimension d.

(a) Deduce the leading asymptotic (8.22).

(b) Show that the strength of the source varies according to Eq. (8.24) in the long-time limit.

(c) Complete the derivation of Eq. (8.26). You may need to compute integrals of the form

$$\int_0^\infty e^{-x} I_0(x)\, dx, \quad \int_0^\infty e^{-x} I_1(x)\, dx, \quad \int_0^\infty e^{-2x} I_0(x) I_1(x)\, dx.$$

For these computations, you may find the following integral representations helpful:

$$I_0(x) = \frac{1}{2\pi} \int_0^{2\pi} e^{x\cos q}\, dq, \quad I_1(x) = \frac{1}{2\pi} \int_0^{2\pi} \cos q\, e^{x\cos q}\, dq.$$

8.3 Investigate Ising–Glauber dynamics in the mean-field approximation. Let the number of spins be finite but very large.

(a) For zero-temperature dynamics determine $P_M(t)$, the probability to have M up spins and $N - M$ down spins.

(b) Determine $P_M(t)$ for critical dynamics.

(c) In the low-temperature regime ($\beta_c < \beta < \infty$), the distribution $P_M(t)$ is bimodal with peaks of width of the order of $\sqrt{N}$ around $M_\pm = \frac{1}{2} N (1 \pm m_\infty)$. The system spends almost all the time near one of the peaks but occasionally makes transitions from one peak to the other. Estimate the transition time.

8.4 Consider Ising–Glauber dynamics at zero temperature in one dimension with the antiferromagnetic initial state.

(a) Compute the pair correlation function $G_k(t)$.

(b) Verify that the domain-wall density is given by (8.59).

(c) Show that the density of domain-wall doublets is given by $\rho_2 = e^{-2t}[I_0(2t) - I_1(2t)]$. Verify that its asymptotic behavior is the same as that for the uncorrelated initial condition, Eq. (8.60).

8.5 Consider the one-dimensional Ising–Glauber model at zero temperature starting from the uncorrelated initial condition (with m_0).

(a) Write evolution equations for the three-spin correlation function $\langle s_i s_j s_k \rangle$, where $i \le j \le k$ but otherwise arbitrary.

(b) Verify (here the initial condition is important) that the solution to these equations is trivial: $\langle s_i s_j s_k \rangle = 0$.

(c) Write down evolution equations for the four-spin correlation function $\langle s_i s_j s_k s_\ell \rangle$, where $i \le j \le k \le \ell$.

(d) Verify that the four-spin correlation function can be expressed via the pair correlation functions:

$$\langle s_i s_j s_k s_\ell \rangle = \langle s_i s_j \rangle \langle s_k s_\ell \rangle + \langle s_i s_\ell \rangle \langle s_j s_k \rangle - \langle s_i s_k \rangle \langle s_j s_\ell \rangle.$$

(e) Show that the previous formula reduces to (8.62) for four consecutive spins.

8.6 Show that the density of domain walls ρ and the density of domain-wall doublets ρ_2 are related by $\dot{\rho} = -2\rho_2$. Verify that the densities (8.58) and (8.60) agree with $\dot{\rho} = -2\rho_2$.

8.7 Compute the asymptotic behavior of the density of domain-wall triplets defined in Eq. (8.61). Explain heuristically why both ρ_3 and $(\rho_2)^2$ decay as t^{-3}.

8.8 Suppose that the Ising chain is endowed with the dynamics of Eq. (8.45) with parameters $\gamma = 2, \delta = 1$.
   (a) Show that this dynamics corresponds to zero temperature.
   (b) Verify that both energy-increasing and energy-conserving spin flips are forbidden for this dynamics.
   (c) Show that the system falls into a jammed state where the distances between adjacent domain walls is greater than the lattice spacing.
   (d) Map the spin evolution onto a random sequential adsorption process. Show that the antiferromagnetic initial state maps onto an empty system, so the final density of domain walls can be read off the results from Chapter 7.
   (e) Compute the final density of domain walls in the case when the initial state is uncorrelated.

8.9 Consider the Ising chain in a non-zero magnetic field with zero-temperature Glauber dynamics. Solve for the density of down domains of length n, for the initial condition of alternating domains of up and down spins that are all of length L.

8.10 Starting with the master equation (8.72), check that the total density of domains $\rho = \sum_k P_k$ satisfies (8.70) and that the total domain length is conserved, $\sum_k k\, dP_k/dt = 0$.

8.11 Solve for the domain number distribution in the Potts model with zero-temperature Glauber dynamics. *Hint*: P_1 is replaced by $P_1/(q-1)$ in Eq. (8.72).

8.12 For the Ising model on the square lattice at a positive temperature, show that the spin-flip rate can be written in a form similar to (8.78), with coefficients

$$\frac{1}{4}\tanh(2\beta J) + \frac{1}{8}\tanh(4\beta J) \qquad \text{and} \qquad \frac{1}{4}\tanh(2\beta J) - \frac{1}{8}\tanh(4\beta J)$$

instead of 3/8 and 1/8.

8.13 Generalize the derivation for Eq. (8.78) and derive the spin-flip rate for the Ising model on the cubic lattice $(d = 3)$ with zero-temperature Glauber dynamics.

8.14 Consider the Ising model on any odd-coordinated lattice with zero-temperature Glauber dynamics. Show that, for any sufficiently random initial condition, e.g. for an uncorrelated initial condition with arbitrary magnetization m_0, an infinite system will never reach a ground state if it is not initially in the ground state.

8.15 Show that, for the Ising model on an $L \times L$ square lattice with periodic boundary conditions, the total number of stripe states is

$$\frac{g^{L-1} - (-g)^{-L+1}}{\sqrt{5}} - \frac{2}{\sqrt{3}}\sin\frac{\pi(L-1)}{3},$$

where $g = (\sqrt{5} + 1)/2$ is the golden ratio.

8.16 Study the Ising chain with zero-temperature Kawasaki dynamics.
   (a) Write the evolution equation for the mean spin at site i, S_i.
   (b) Show that the correlation functions obey an infinite hierarchy of equations.
   (c) Solve for the domain-wall density for random initial conditions. Compare your result with that of Eq. (8.91) for the antiferromagnetic initial condition.
   (d) Obtain the number of frozen configurations.

8.17 Consider the Ising chain with very low-temperature Kawasaki dynamics. We want to compute the probability $E(x)$ that a spin that is at position x inside a domain of length L will ultimately reach $x = L$.
   (a) Determine the equations that are obeyed by the exit probability, paying particular attention to the cases $x = 1$ and $x = L - 1$.
   (b) Solve these equations and show that $E(x) = (2x - 1)/(2L - 2)$ for all $0 < x < L$.

8.18 Consider the zero-temperature Swendsen–Wang dynamics for the $q = \infty$ Potts model, in which only energy-lowering events can occur. In this case, the flipping of a domain causes it to merge with only one of its neighbors. Determine the density of domains and the domain length distribution.

Coarsening

In the previous chapter, we discussed the phase-separation kinetics or domain coarsening in the kinetic Ising model after a temperature quench from a homogeneous high-temperature phase to a two-phase low-temperature regime. Because of the complexity of the ensuing coarsening process, considerable effort has been devoted to developing continuum, and analytically more tractable, theories of coarsening. While a direct connection to individual spins is lost in such a continuum formulation, the continuum approach provides many new insights that are hard to obtain by a description at the level of individual spins.

9.1 Models

We tacitly assume that the order parameter is a scalar unless stated otherwise. We generally have in mind magnetic systems and will use the terminology of such systems; this usage reflects tradition rather than the dominant application of coarsening. There is a crucial distinction between non-conservative and conservative dynamics, and we begin by describing generic models for these two dynamics.

Non-conservative dynamics

The basic ingredients that underlie non-conservative dynamics are the following.

- The primary variable is a continuous coarse-grained order parameter, or magnetization $m(\mathbf{x}, t) \equiv \ell^{-d} \sum \sigma$, the average magnetization in a block of linear dimension ℓ that is centered at $\mathbf{x}$, rather than a binary Ising variable $\sigma = \pm 1$. Here ℓ should be much greater than the lattice spacing a and much smaller than the system size to give a smoothly varying coarse-grained magnetization on a scale greater than ℓ. This coarse graining applies over a time range where the typical domain size remains small compared to the system size.
- The thermodynamics of the system are described by the coarse-grained Landau free energy functional

$$F[m(\mathbf{x})] = \int \left[\frac{1}{2} \left| \nabla m(\mathbf{x}) \right|^2 + V(m(\mathbf{x})) \right] d\mathbf{x}, \qquad (9.1)$$

where the potential $V(m)$ has a double-well structure whose minima correspond to the equilibrium states. The standard example is $V(m) = \frac{1}{2} (1 - m^2)^2$ for systems with two degenerate minima.

- The final step to construct a coarse-grained description is also phenomenological: we assert that the order parameter changes at a rate that is proportional to the local thermo-dynamic force $\delta F / \delta m(\mathbf{x}, t)$, the functional derivative of F with respect to m at position $\mathbf{x}$. We are considering *overdamped* dynamics because there is no second time derivative in the equation of motion. This assumption is equivalent to ignoring inertia. By absorb-ing a coefficient of proportionality into the time scale, we arrive at the *time-dependent Ginzburg–Landau* (TDGL) equation

$$\frac{\partial m}{\partial t} = -\frac{\delta F}{\delta m} = \nabla^2 m - V'(m), \tag{9.2}$$

which is one of the central equations of coarsening.

There are several important aspects of the TDGL equation that we wish to emphasize.

1. While this coarse-grained free energy (9.1) is intuitively plausible – it combines the Landau mean-field theory with the lowest-order contribution due to spatial variation of the order parameter – it is not possible to derive Eq. (9.1) and the TDGL equation from first principles by starting with a microscopic spin model.
2. The TDGL equation is purely dissipative; that is, the change in the free energy as a function of time is

$$\frac{dF}{dt} = \int \frac{\delta F}{\delta m} \frac{\partial m}{\partial t} \, d\mathbf{x} = -\int \left(\frac{\delta F}{\delta m}\right)^2 \, d\mathbf{x} \le 0.$$

Thus a system governed by the TDGL equation "flows" down the free energy gradient until a potential minimum is reached.

3. The TDGL equation (9.2) applies to zero temperature. At a positive temperature, Eq. (9.2) must be supplemented by a noise term whose intensity grows with temperature. Thus more generally the TDGL equation is a stochastic nonlinear partial differential equation. Below the critical temperature, the dynamical behavior is essentially universal (this is part of the central dogma mentioned at the outset of Chapter 8) and therefore we can focus on zero-temperature dynamics.
4. The TDGL equation is deterministic at zero temperature. In contrast, the Ising–Glauber description remains stochastic at zero temperature.
5. The TDGL equation can be interpreted as a reaction–diffusion process. For example, for the potential $V(m) = \frac{1}{2}(1 - m^2)^2$, the TDGL equation is:

$$\frac{\partial m}{\partial t} = \nabla^2 m + 2m(1 - m^2). \tag{9.3}$$

If we view m as a density, then the above equation describes the evolution of a diffusing population of particles that give birth, $A \to 2A$, and undergo three-body coalescence, $3A \to A$ (alternatively, the loss process could be annihilation $3A \to \emptyset$). For this reaction, the density evolves according to $dm/dt = 2m(1 - m^2)$, which has an unstable fixed point at $m = 0$ and a stable fixed point at $m = 1$. When diffusion is included, the resulting equation of motion (9.3) describes the infiltration of a stable high-density phase ($m = 1$)

into a low-density ($m = 0$) region. In fact, Eq. (9.3) is one example[1] from a family of self-regulating reactions of the form

$$\frac{\partial m}{\partial t} = \nabla^2 m + f(m), \tag{9.4}$$

with $f(0) = f(1) = 0$, and $f(m)$ having a single maximum in the range $[0, 1]$. This interpretation involves a caveat, however. For reaction–diffusion processes, m represents a density which must be positive; however, for the TDGL, the scalar order parameter m can be negative. Mathematically, however, the equivalence to reaction–diffusion processes is useful as an intuitive guide to the resulting phenomenology.

Conservative dynamics

In two-component alloy systems, the natural order parameter is the difference in the concentration of the constituent elements. By construction, this order parameter is conserved in an isolated piece of material. Thus a dynamics different from the TDGL equation is needed to account for this conservation. At a phenomenological level, we seek a governing equation that ensures that the flux of atoms of each element of the alloy can be expressed as a (conserved) continuity equation

$$\frac{\partial m}{\partial t} + \nabla \cdot \mathbf{J} = 0, \tag{9.5}$$

where we again consider overdamped dynamics. In Eq. (9.5), the flux vector $\mathbf{J}$ should depend on the order parameter through the free energy (9.1). The simplest choice that is both conservative and involves gradient flow is $\mathbf{J} \sim -\nabla \delta F / \delta m$. We again absorb the proportionality factor into the time scale and we arrive at the *Cahn–Hilliard* (CH) equation

$$\frac{\partial m}{\partial t} = \nabla^2 \frac{\delta F}{\delta m} = -\nabla^2 \left[\nabla^2 m - V'(m) \right], \tag{9.6}$$

which plays the same role for zero-temperature conservative dynamics as the TDGL equation plays for non-conservative dynamics.

In forms given in Eqs (9.2) and (9.6), the governing TDGL and CH equations, are deterministic; this means that we are dealing with systems at zero temperature. To phenomenologically model the behavior at a positive temperature, a (Langevin) noise term should be added to the right-hand side of these two equations. As articulated in the central dogma at the outset of the previous chapter, this additional term should not change qualitative dynamical behavior as long as the temperature is below the critical temperature T_c.

Typically, one starts with the initial temperature $T_i = \infty$, where the system is completely disordered, and quenches the system to $T_f = 0$. The coarse-grained form of this disordered initial condition is

$$\langle m(\mathbf{x}, 0) \rangle = 0, \quad \langle m(\mathbf{x}, 0) \, m(\mathbf{x}', 0) \rangle = \delta(\mathbf{x} - \mathbf{x}'). \tag{9.7}$$

---

[1] The most famous such example, the Fisher–Kolmogorov–Petrovsky–Piscounov (FKPP) equation, arises for $f(m) = m(1 - m)$. This example will be discussed in Chapter 13.

The goal is to determine the long-time behavior of the solutions of the deterministic TDGL and CH equations subject to these random initial conditions. It is now well established that a scale-invariant coarsening domain mosaic emerges from the solutions to these equations at late times. This morphology is (statistically) independent of time when all lengths are rescaled by a typical domain size $L(t)$ that grows algebraically with time. This characteristic length scale is usually written as $L(t) \sim t^z$, with z the dynamical exponent whose value is given by (as in (8.94))

$$z = \begin{cases} 1/2 & \text{(TDGL)}, \\ 1/3 & \text{(CH)}, \end{cases} \tag{9.8}$$

for scalar order-parameter systems. The existence of this universal dynamical exponent – in which the conservation or non-conservation of the order parameter is the only relevant feature – is one of the tenets of the central dogma.

9.2 Free evolution

We now use scaling to simplify the description of the dynamics. The evidence in favor of scaling is compelling, but it has not been proved except for the TDGL equation in one dimension and for special microscopic models, such as the one-dimensional Ising–Glauber model and the n-vector model with $n = \infty$. It is therefore reassuring to see the emergence of scaling in the TDGL and the CH equations, even if these examples are idealized. In this spirit, we now study the case where the potential vanishes, $V(m) = 0$. The dynamical equation then reduces to the diffusion equation, $\partial m / \partial t = \nabla^2 m$, for non-conserved order-parameter dynamics, and to the bidiffusion equation, $\partial m / \partial t = -\nabla^4 m$, for conserved order-parameter dynamics. In these two examples, dimensional analysis suggests the existence of a single growing length scale, $L(t) \sim t^{1/2}$ (TDGL) and $L(t) \sim t^{1/4}$ (CH).[2]

If a spin system is initially in the random state given by Eq. (9.7), then the global magnetization remains zero throughout the evolution. We therefore need two-body correlation functions to investigate the nature of coarsening; we already encountered this same issue in our study of the voter model and the Ising–Glauber model in the previous chapter. We thus define the two-body correlation function

$$C(\mathbf{r}_1, t_1, \mathbf{r}_2, t_2) \equiv \frac{\langle m(\mathbf{r}_1, t_1)\, m(\mathbf{r}_2, t_2) \rangle}{\sqrt{\langle m^2(\mathbf{r}_1, t_1) \rangle\, \langle m^2(\mathbf{r}_2, t_2) \rangle}} \tag{9.9}$$

to probe the domain structure at two different space-time points. We have normalized the correlation function so as to make it dimensionless. For brevity, we use the shorthand notation $1 \equiv (\mathbf{r}_1, t_1)$ and $2 \equiv (\mathbf{r}_2, t_2)$. Translational invariance implies that $C(1, 2) = C(\mathbf{r}, t_1, t_2)$ with $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$. It is also useful to explicitly study the *autocorrelation function*

---

[2] See Section 1.1 for additional discussion on dimensional analysis and scaling.

$A(t_1, t_2) = C(\mathbf{0}, t_1, t_2)$ that measures the probability that the sign of the magnetization coincides at times t_1 and t_2.

Let's begin by investigating coarsening in the potential-free TDGL (diffusion) equation

$$\frac{\partial m}{\partial t} = \nabla^2 m \tag{9.10}$$

for the random initial condition of Eq. (9.7). The solution is

$$m(\mathbf{r}_1, t) = \frac{1}{(4\pi t)^{d/2}} \int m(\mathbf{z}_1, 0)\, e^{-(\mathbf{r}_1 - \mathbf{z}_1)^2/4t}\, d\mathbf{z}_1\,.$$

The average of the product of the magnetization at two different space-time points is

$$\langle m(\mathbf{r}_1, t_1) m(\mathbf{r}_2, t_2) \rangle = \frac{1}{[(4\pi)^2 t_1 t_2]^{d/2}} \iint \underbrace{\langle m(\mathbf{z}_1, 0)\, m(\mathbf{z}_2, 0) \rangle}_{\delta(\mathbf{z}_1 - \mathbf{z}_2)}$$

$$\times\, e^{-(\mathbf{r}_1 - \mathbf{z}_1)^2/4t_1}\, e^{-(\mathbf{r}_2 - \mathbf{z}_2)^2/4t_2}\, d\mathbf{z}_1 d\mathbf{z}_2$$

$$= \frac{1}{[(4\pi)^2 t_1 t_2]^{d/2}} \int e^{-(\mathbf{r}_1 - \mathbf{z})^2/4t_1}\, e^{-(\mathbf{r}_2 - \mathbf{z})^2/4t_2}\, d\mathbf{z}$$

$$= \frac{1}{[(4\pi)^2 t_1 t_2]^{d/2}} \int e^{-\mathbf{r}_1^2/4t_1}\, e^{-\mathbf{r}_2^2/4t_2}\, e^{-\mathbf{z}^2(1/4t_1 + 1/4t_2)}$$

$$\times\, e^{\mathbf{z}\cdot(\mathbf{r}_1/2t_1 + \mathbf{r}_2/2t_2)}\,.$$

We perform the integral in the last line by completing the square in the exponential to give

$$\langle m(\mathbf{r}_1, t_1) m(\mathbf{r}_2, t_2) \rangle = \frac{1}{[4\pi(t_1 + t_2)]^{d/2}}\, e^{-(\mathbf{r}_1 - \mathbf{r}_2)^2/4(t_1 + t_2)}\,. \tag{9.11}$$

Finally, the normalized two-body correlation function defined by Eq. (9.9) is

$$C(1, 2) = \left[\frac{2\sqrt{t_1 t_2}}{t_1 + t_2}\right]^{d/4}\, e^{-(\mathbf{r}_1 - \mathbf{r}_2)^2/4(t_1 + t_2)}\,. \tag{9.12}$$

Thus the correlation function decays quickly (as a Gaussian) as a function of spatial separation, but slowly (as a power law) as a function of temporal separation.

The potential-free CH (bidiffusion) equation

$$\frac{\partial m}{\partial t} = -\nabla^4 m \tag{9.13}$$

also admits a full analytical treatment of coarsening. The simplest way to solve the bidiffusion (and the diffusion) equations is with Fourier transforms. Using the Fourier expansion $m(\mathbf{r}, t) = (2\pi)^{-d} \int m(\mathbf{k}, t)\, e^{-i\mathbf{k}\cdot\mathbf{r}}\, d\mathbf{k}$, we reduce (9.13) to the ordinary differential equation $dm/dt = -k^4 t$, whose solution is $m(\mathbf{k}, t) = m_0(\mathbf{k})\, e^{-k^4 t}$, with $m_0(\mathbf{k}) \equiv m(\mathbf{k}, t = 0)$. Inverting this Fourier transform gives

$$m(\mathbf{r}, t) = (2\pi)^{-d} \int m_0(\mathbf{k})\, e^{-i\mathbf{k}\cdot\mathbf{r} - k^4 t}\, d\mathbf{k}. \tag{9.14}$$

A more useful quantity is the mean-square magnetization because the stochasticity of the initial condition is averaged out, just as in the classic Langevin equation (Section 2.8). For the white-noise initial condition (9.7), the Fourier transform and the correlations between Fourier components are

$$\langle m_0(\mathbf{k})\rangle = 0, \quad \langle m_0(\mathbf{k})\, m_0(\mathbf{k}')\rangle = (2\pi)^d \delta(\mathbf{k}+\mathbf{k}'). \tag{9.15}$$

Using these results along with Eqs (9.7) and (9.14), the mean-square magnetization $\langle m^2(\mathbf{r},t)\rangle$ is given by

$$\langle m^2(\mathbf{r},t)\rangle = (2\pi)^{-d}\int e^{-2k^4 t}\, d\mathbf{k} = B_d (2t)^{-d/4}, \tag{9.16}$$

where $B_d = (2\pi)^{-d}\Omega_d \Gamma(d/4)/4$. Here $\Omega_d = 2\pi^{d/2}/\Gamma(d/2)$ is the area of the unit sphere in d dimensions and Γ is the Euler gamma function.

Similarly, the correlation function is

$$\langle m(1)\, m(2)\rangle = (2\pi)^{-d}\int e^{-i\mathbf{k}\cdot\mathbf{r}-k^4(t_1+t_2)}\, d\mathbf{k}. \tag{9.17}$$

When spatial points coincide, $\mathbf{r}=\mathbf{0}$, we obtain the autocorrelation function

$$A(t_1,t_2) = \left[\frac{2\sqrt{t_1 t_2}}{t_1+t_2}\right]^{d/4}, \tag{9.18}$$

which has the same form for both the diffusion and the bidiffusion equation. If the spatial points are different, $\mathbf{r}_1 \neq \mathbf{r}_2$, we perform the integral in Eq. (9.17) by introducing spherical coordinates in $\mathbf{k}$-space for $d \geq 2$. Then $\mathbf{k}\cdot\mathbf{r} = kr\cos\theta$ and $d\mathbf{k} = k^{d-1}\Omega_{d-1}\sin^{d-2}\theta\, d\theta\, dk$, and we obtain

$$\langle m(1)\, m(2)\rangle = \frac{\Omega_{d-1}}{(2\pi)^d}\int_0^\infty k^{d-1}\, e^{-k^4(t_1+t_2)}\, F_d(kr)\, dk,$$

where

$$F_d(u) = \int_0^\pi (\sin\theta)^{d-2}\, e^{-iu\cos\theta}\, d\theta.$$

Computing this integral we obtain the two-body correlation function $C(r,t_1,t_2)$:

$$C(r,t_1,t_2) = \begin{cases} \dfrac{2(4t_1 t_2)^{1/8}}{\Gamma(1/4)\,r}\displaystyle\int_{-\infty}^\infty \cos q\, e^{-q^4\tau}\, dq, & d=1, \\[2ex] \dfrac{4(4t_1 t_2)^{1/4}}{\sqrt{\pi}\,r^2}\displaystyle\int_0^\infty q\, I_0(q)\, e^{-q^4\tau}\, dq, & d=2, \\[2ex] \dfrac{2(4t_1 t_2)^{3/8}}{\Gamma(3/4)\,r^3}\displaystyle\int_0^\infty q\sin q\, e^{-q^4\tau}\, dq, & d=3. \end{cases} \tag{9.19}$$

Here, we use the shorthand notations $q = kr$, $\tau = (t_1 + t_2)/r^4$, and I_0 is the modified Bessel function of order zero.

Finally, the *temporally normalized* correlation function $G(\mathbf{r}, t_1, t_2) \equiv C(\mathbf{r}, t_1, t_2)/A(t_1, t_2)$ is a function of a *single* scaling variable τ (here "temporally normalized" reflects the property that $G(\mathbf{0}, t_1, t_2) \equiv 1$). This single-parameter scaling is a peculiarity of the toy model rather than a universal rule; generally, rotational symmetry and dynamical scaling would imply that $G(\mathbf{r}, t_1, t_2)$ is a function of two variables. If $t_1 = t_2 = t$, then both the correlation functions $C(\mathbf{r}, t, t)$ and $G(\mathbf{r}, t, t)$ reduce to the equal-time correlation function $G(\tau)$, which is a function of the single scaling variable $\tau = 2t/r^4$. For the potential-free CH equation, the precise form of the equal-time correlation function $G(\tau)$ is

$$
G(\tau) = \begin{cases}
\dfrac{2\tau^{1/4}}{\Gamma(1/4)} \displaystyle\int_{-\infty}^{\infty} \cos q \; e^{-q^4 \tau} \, dq, & d = 1, \\[2ex]
\dfrac{4\tau^{1/2}}{\sqrt{\pi}} \displaystyle\int_{0}^{\infty} q \, I_0(q) \; e^{-q^4 \tau} \, dq, & d = 2, \\[2ex]
\dfrac{2\tau^{3/4}}{\Gamma(3/4)} \displaystyle\int_{0}^{\infty} q \, \sin q \; e^{-q^4 \tau} \, dq, & d = 3.
\end{cases}
$$

In the absence of a potential, the dynamics of the order parameter and its correlation functions are completely soluble. In addition to determining the growth of the typical length scale, more detailed properties are accessible, such as the decay of the autocorrelation function. This solubility provides a strong motivation for studying these free-field models. Additionally, the mechanism that gives the different dynamical exponent for the diffusion and bidiffusion equations is apparent. One must nevertheless be aware of the limitations of the free-field models; for example, the time dependence obtained by analysis of the bidiffusion equation, $L(t) \sim t^{1/4}$, differs from the well-established $t^{1/3}$ growth that typifies coarsening processes with a conserved scalar order parameter.[3]

9.3 Case studies in non-conservative dynamics

We now turn to explicit solutions of the equations of motion. Because domain interfaces in coarsening are geometrically complex, we first study idealized settings that admit an exact analysis but still are dynamically non-trivial. There are several classical examples of this kind, such as a single domain wall and a single spherical droplet. We also discuss the finger and wedge geometries that shed much more light on the dynamics of single interfaces.

Before looking at spatially non-trivial configurations, let's verify that homogeneous configurations evolve to the ground states. For a spatially homogeneous system, the TDGL equation (9.2) simplifies to $dm/dt = -V'(m)$ and therefore the system indeed reaches one of the minima of $V(m)$ in the long-time limit. In what follows, we consider the double-well

---

[3] However, $t^{1/4}$ growth does arise for conserved systems with a vector order parameter (with some exceptions for the *XY* model).

potential $V(m) = \frac{1}{2}(1 - m^2)^2$ if not stated otherwise. For this potential, the spatially homogeneous and stationary solution is either $m = 1$ or $m = -1$. Our goal is to understand the evolution for spatially inhomogeneous systems.

Straight domain wall

The simplest example of spatial inhomogeneity is a single interface in the form of a straight domain wall. For the discrete Ising–Glauber model, the magnetization is merely a step function. For the continuum TDGL equation, a flat interface is a bit more interesting because the coarse-grained magnetization has a non-trivial spatial variation across the interface. For a flat stationary interface, the TDGL equation reduces to the ordinary differential equation

$$\frac{d^2m}{dx^2} = V'(m), \tag{9.20}$$
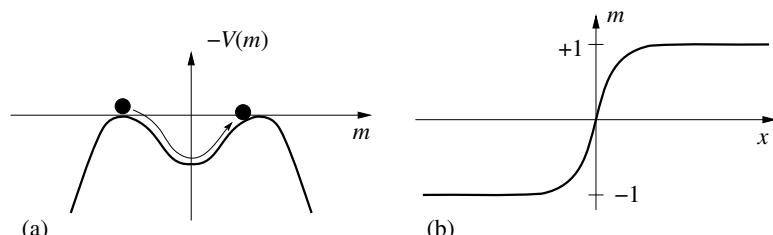
where x is the coordinate along the axis orthogonal to the interface. Far from the interface the order parameter is homogeneous and stationary; for the generic double-well potential, the boundary condition for (9.20) is $m(\pm\infty) = \pm 1$.

Before solving (9.20), we outline an appealing interpretation of the domain wall in terms of classical mechanics. By treating x as the time variable and m as the coordinate, we see that a straight domain wall is equivalent to the motion of a fictitious particle of unit mass that moves in the potential $-V$ (Fig. 9.1). The energy of this fictitious particle is

$$E = \frac{1}{2}\left(\frac{dm}{dx}\right)^2 - V(m), \tag{9.21}$$

and does not change with "time" x. The boundary conditions imply that $E = 0$. Thus $dm/dx = \sqrt{2V(m)}$, or

$$x = \int_{m(0)}^{m(x)} \frac{d\mu}{\sqrt{2V(\mu)}}. \tag{9.22}$$

(a)   (b)

Fig. 9.1.   (a) The effective potential $-V(m) = -\frac{1}{2}(1 - m^2)^2$. Also illustrated is the motion of the fictitious particle that corresponds to a single kink. This particle starts at the top of the left maximum at $m = -1$ and moves without dissipation until it eventually stops at the maximum at $m = +1$. (b) The corresponding dependence of m on x.

It is customary to set the "initial condition" to be $m(x = 0) = 0$. For the potential $V(m) = \frac{1}{2}(1 - m^2)^2$, the solution to Eq. (9.22) gives the domain-wall profile

$$m(x) = \tanh x . \tag{9.23}$$

This solution is known as a "kink" in a ϕ^4 theory. The width of this kink is of the order of 1, while its free energy density is (see (9.1))

$$\frac{1}{2}\left(\frac{dm}{dx}\right)^2 + V(m) = 2V(m(x)) = \frac{1}{\cosh^4 x} , \tag{9.24}$$

and the total free energy per unit area, that is, the surface tension, is

$$\sigma = \int_{-\infty}^{\infty} 2V(m)\, dx = \int_{-\infty}^{\infty} 2V(m)\left(\frac{dm}{dx}\right)^{-1} dm = \int_{-1}^{1} \sqrt{2V(m)}\, dm = \frac{4}{3} . \tag{9.25}$$

Within this classical mechanical analogy, the governing equation (9.20) can be derived by minimizing the "action"

$$\int_{-\infty}^{\infty}\left[\frac{1}{2}\left(\frac{dm}{dx}\right)^2 + V(m)\right] dx.$$

Because the kinetic and potential energies are equal, the action associated with the fictitious particle coincides with the surface tension.

The mechanical analogy provides many useful insights about the phase behavior of a system, especially in situations where an analytical description of domain walls is not feasible. For example, suppose that the initial condition is a combination of two static solutions, such as a kink–antikink configuration. Can such a configuration be stable? Within the mechanical analogy, we are asking if it is possible for the fictitious particle to start at the maximum at $m = -1$, move to the maximum at $m = +1$, and then return to the maximum at $m = -1$ (see Fig. 9.1). Clearly this motion is impossible, so we may conclude without calculation that a kink–antikink pair cannot be stable. Numerical simulations show that the kink and antikink slowly move toward each other and eventually annihilate. We shall analyze this motion later in this chapter.

Another useful aspect of the mechanical analogy is that we can use it to characterize *all* possible one-dimensional stationary solutions that correspond to a finite free energy per unit area. For any potential function, (i) the fictitious particle must start at one maximum of the potential $[-V]$ at "time" $x = -\infty$ and reach another maximum of the potential $[-V]$ at $x = \infty$, and (ii) the maxima must be adjacent. Thus for potentials with two degenerate minima (at $m = \pm 1$) there are two types of static solutions – a kink and an antikink – that we symbolically denote as $[-1, 1]$ and $[1, -1]$, respectively. Similarly for potentials with three degenerate minima, for example, a three-state spin system with minima at $m = \pm 1$ and $m = 0$ (as exemplified by the potential $V(m) = \frac{1}{2}m^2(1 - m^2)^2$), there are four types of static solutions: $[-1, 0]$, $[0, -1]$, $[1, 0]$, and $[0, 1]$. The exact static solutions of Eq. (9.22) that separate the $m = 0$ and $m = \pm 1$ phases for this potential are given by (problem 9.3)

$$m(x) = \pm\left(1 + e^{\pm x}\right)^{-1/2} . \tag{9.26}$$

Moreover, it is obvious from the mechanical analogy that the fictitious particle cannot start at the maximum at $m = -1$ and move to the maximum at $m = +1$; thus a domain wall of the form $[-1, 1]$ cannot occur.

Stability of domain walls and ground states

In one dimension, the TDGL equation admits spatially varying time-independent solutions (domain walls) if the system has at least two ground states. Physically, the stability of ground states and domain walls is evident. The TDGL equation has gradient dynamics, so that the free energy decreases and eventually reaches a minimum. Since domain walls and ground states are the only possible minima, the system should fall into one of these states. A potential pitfall in this argument is the tacit assumption of the existence of a finite number of domain walls, such as a kink and an antikink in the case where the potential has two degenerate minima. The number of *types* of domain walls is indeed finite; however, due to translational invariance, there are infinitely many domain walls of a given type. Thus if we start with a domain wall and perturb it a bit, we anticipate that the magnetization will eventually relax to a domain wall of the same type but shifted by a certain distance.

To determine the validity of this picture and to make it more quantitative, let us study a small perturbation about a stationary domain wall. We therefore write the magnetization as the sum of the stationary solution $m_0 = \tanh x$ plus a small perturbation $\phi(x, t)$:

$$m(x, t) = m_0(x) + \phi(x, t). \tag{9.27}$$

Substituting this ansatz into the TDGL equation (9.2) and expanding, we obtain (to leading order)

$$\frac{\partial \phi}{\partial t} = \frac{\partial^2 \phi}{\partial x^2} - V''(m_0)\phi. \tag{9.28}$$

This is a time-dependent one-dimensional Schrödinger equation with potential $V''(m_0)$ whose solution is[4]

$$\phi(x, t) = \sum_{n \geq 0} A_n e^{-E_n t} \psi_n(x), \tag{9.29}$$

where the coefficients A_n are determined from the initial condition, and E_n and ψ_n are respectively the eigenvalues and the eigenfunctions that satisfy

$$E_n \psi_n = -\frac{d^2 \psi_n}{dx^2} + V''(m_0)\psi_n. \tag{9.30}$$

The solution $m_0(x)$ is stable under small perturbations if all the eigenvalues of the Schrödinger equation (9.30) are non-negative. Let us establish this assertion for the potential

---

[4] Equation (9.28) is actually a Schrödinger equation in imaginary time. Note also that the sum in (9.29) is schematic; it includes both a discrete spectrum that contains a few (sometimes just one) elements and an integration over the continuous part of the spectrum.

$V = \frac{1}{2}(1 - m^2)^2$. For a kink that is centered at the origin, $m_0 = \tanh x$, Eq. (9.30) becomes

$$E_n \psi_n = \left[-\frac{d^2}{dx^2} + (6\tanh^2 x - 2) \right] \psi_n . \tag{9.31}$$

This Schrödinger equation appears in many textbooks on quantum mechanics.[5] The energy spectrum consists of a small number of discrete levels followed by a continuous spectrum. Adapting these results to our specific example we arrive at the following outcome: there are two discrete levels

$$E_0 = 0 \quad \text{with} \quad \psi_0 = 1/\cosh^2 x, \tag{9.32a}$$

$$E_1 = 3 \quad \text{with} \quad \psi_1 = \sinh x/\cosh^2 x, \tag{9.32b}$$

and the continuous spectrum for $E \geq 4$. (This last result is to be expected, as the potential $6\tanh^2 x - 2$ in the Schrödinger equation (9.30) tends to 4 as $x \to \pm\infty$, so the continuous spectrum must be at higher energy.)

Some details of the predictions (9.32a)–(9.32b), particularly regarding the first excited state ψ_1, are sensitive to the choice of $V(m)$; other features are generic and can be obtained without calculations. For example, Eq. (9.30) always admits an eigenfunction

$$\psi_0 = \frac{dm_0}{dx} , \tag{9.33}$$

with eigenvalue $E_0 = 0$. Indeed, differentiating the governing TDGL equation

$$\frac{d^2 m_0}{dx^2} = V'(m_0) ,$$

we see that ψ_0 given by (9.33) satisfies (9.30) with $E_0 = 0$. A more conceptual derivation of the same result is to notice that $m_0(x + a)$ is a stationary solution and if we expand in powers of a we obtain

$$m_0(x + a) = m_0(x) + a\psi_0(x)$$

to leading order. Therefore $a\psi_0(x)$ satisfies (9.28), and since this function is time-independent, we find again that ψ_0 is an eigenfunction with $E_0 = 0$.

Shrinking of a droplet

A tractable example of interface dynamics in higher dimensions is the evolution of a droplet of one phase, say $m = -1$, in a sea of the opposite phase, with $m = +1$. From the perspective of individual spins, we expect that a finite cluster of minority spins will be overwhelmed by the large majority of oppositely oriented spins. This same behavior also occurs in the continuum picture.

---

[5] The potential is usually written as $-U_0(\cosh \alpha x)^{-2}$. Choosing $U_0 = 6$ and $\alpha = 1$ we recover (up to a shift in energy) the potential that appears in (9.31).

Since the droplet shrinks, we must use the full time-dependent TDGL equation to describe its evolution. Owing to spherical symmetry, the TDGL equation (9.2) reduces to

$$\frac{\partial m}{\partial t} = \frac{\partial^2 m}{\partial r^2} + \frac{d-1}{r} \frac{\partial m}{\partial r} - V'(m). \tag{9.34}$$

Provided that the droplet radius R greatly exceeds the interface width, we anticipate that the magnetization has the form $m(r, t) = f(r - R(t))$. Substituting this ansatz into (9.34) and defining $x = r - R(t)$ gives

$$f'' + \left(\frac{d-1}{r} + \frac{dR}{dt} \right) f' - V'(f) = 0. \tag{9.35}$$

Let's multiply this equation by f' and integrate with respect to x through the interface. For a large droplet, the interface is localized within a narrow region around $R(t)$. Consequently, we can set the lower limit of the integral, $-R$, to $-\infty$, leading to

$$(f')^2 \Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \left[\frac{d-1}{r} + \frac{dR}{dt} \right] (f')^2 \, dx - V(x) \Big|_{-\infty}^{\infty} = 0. \tag{9.36}$$

Now $f(x)$ changes suddenly from -1 to $+1$ near $x = 0$ (corresponding to $r = R(t)$), so that f' is sharply peaked near the origin. Thus the first term vanishes. Similarly, the last term is zero, since the potential has the same value at $x = \pm\infty$. Finally, $(f')^2$ in the integral is non-zero and positive only close to $x = 0$. Hence we may replace the term in square brackets in the integrand by its value at $r = R$ to give the equation of motion for the interface:

$$\frac{d-1}{R} + \frac{dR}{dt} = 0. \tag{9.37}$$

Solving this equation, we find that the droplet radius shrinks as $R^2 = R_0^2 - 2(d-1)t$. Consequently, the time for a droplet to disappear is $R_0^2/[2(d-1)]$.

The virtue of focusing on the interfacial region is that we obtain a tractable equation for its dynamics, instead of dealing with the unsolvable TDGL equation (9.34). This effective description of interface dynamics applies for an arbitrarily complex domain geometry, provided that the characteristic length scale associated with the interface morphology greatly exceeds the interface width. This construction is particularly useful in two dimensions. Locally an interface has a curvature $1/R$ and therefore the interface can be approximated by a circle of radius R. As a result, the velocity normal to the interface is $v = -1/R$. Generally in d dimensions, the interface is a $(d-1)$-dimensional manifold. Denoting its principal radii of curvature by $R_1, \ldots, R_{d-1}$ (which depend on the local position on the interface), then the normal interface velocity is the sum of the principal curvatures $1/R_j$; this leads to the *Allen–Cahn* equation

$$v = -(d-1)K, \tag{9.38}$$

where K is the *mean curvature*, $K = (d-1)^{-1} \sum 1/R_j$. There are several points about the Allen–Cahn equation that are worth highlighting:

1. The dynamics is purely *local*; namely, the velocity is determined by the local curvature and is independent of the rest of the interface, the presence of different interfaces, etc. This locality characterizes the TDGL in the asymptotic regime and does *not* hold for conservative dynamics.
2. The dynamics is *independent* of the precise form of the potential $V(m)$.
3. The Allen–Cahn equation does not apply in one dimension, where the interface between two domains is a single point. Here the interface velocity is an exponentially small function of the domain lengths on either side of the interface. This feature leads to ultra-slow logarithmic domain growth that is described by extremal dynamics (see Section 9.7).

Interface dynamics that is derived from the TDGL equation is an example of *curvature-driven* flow, in which arbitrary shapes are driven by the local curvature. Much is known about the evolution of a single closed interface by curvature-driven flows. In two dimensions, every such interface asymptotically approaches a (shrinking) circular shape, independent of its initial shape. It is surprising, at first sight, that a curve of the form given in Fig. 9.2 asymptotically shrinks to a circle by curvature-driven flow, without any pinching off of the most prominent protrusions. In three dimensions, if the initial interface is everywhere *convex*, then the outcome of curvature-driven flow at long times is a sphere. However, an interface that contains both convex and concave portions can undergo fission if the concave portion of the interface is sufficiently thin, such as an elongated liquid drop with a thin neck of fluid in the middle. The classification of all possible topology changes in higher dimensions by curvature-driven flow is still incomplete.

Two-dimensional unbounded domains

In two dimensions, the evolution of a single *closed* interface whose characteristic size greatly exceeds its width is simple in the long-time limit – the interface becomes circular and shrinks according to Eq. (9.37). What about unbounded interfaces? It is again the case
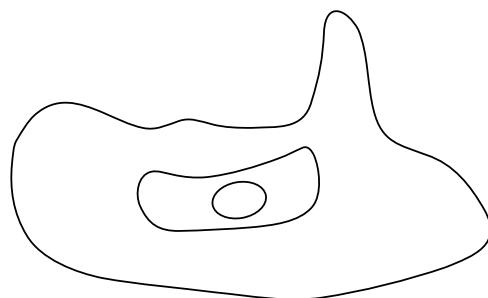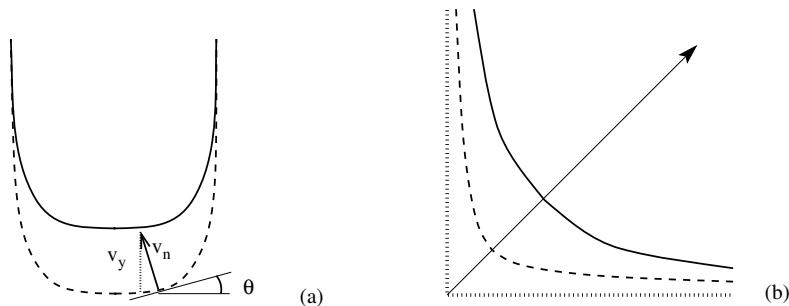
Fig. 9.2. A closed two-dimensional interface that evolves by curvature-driven flow necessarily becomes progressively more circular as it shrinks.

Fig. 9.3. Illustrations of the receding finger (a), showing the normal velocity v_n and the vertical velocity v_y, and the wedge (b), showing the initial wedge (dotted) and the geometry at progressively later times (dashed and solid, respectively).

that a sufficiently simple initial geometry evolves to a stationary state whose properties can be calculated. Here we consider the soluble examples of a semi-infinite finger and an infinite wedge (Fig. 9.3); the latter is the continuous counterpart of the single-corner system that was studied in Section 8.6.

For the finger, the minority phase initially occupies the semi-infinite region $y > 0$ and $|x| < L$. The interesting limit is $t \gg L^2$ where sufficient time has elapsed so that the two corners of the initial finger interact. In this long-time regime, the finger relaxes to a limiting shape that recedes at a constant velocity, so that the interface $y(x)$ is stationary in a co-moving reference frame. For this geometry, the Allen–Cahn equation is

$$v_n = \frac{y''}{[1 + (y')^2]^{3/2}} , \tag{9.39}$$

where v_n is the velocity normal to the interface. The right-hand side is just the curvature of the locus $y(x)$ that defines the finger boundary, and the prime denotes differentiation with respect to x. The vertical component of the velocity is given by $v_y = v_n / \cos\theta = v_n \sqrt{1 + (y')^2}$ (Fig. 9.3). In the steady state, the vertical velocity is constant, $v_y \equiv v$, and hence $y(x)$ satisfies

$$v = \frac{y''}{1 + (y')^2} . \tag{9.40}$$

This equation can be integrated by elementary methods; the first step is to solve for $z \equiv y'$, which satisfies $z' = v(1 + z^2)$, and then integrate a second time. The solution that satisfies the boundary condition $y \to \infty$ when $|x| \to L$ is

$$y = -\frac{2L}{\pi} \ln\left[\cos\left(\frac{\pi x}{2L}\right)\right], \qquad v = \frac{\pi}{2L} . \tag{9.41}$$

The example of the right-angle infinite wedge is a bit more involved. For this example, we let the initial corner be at the origin, so that the wedge occupies the region $x, y \geq 0$. The

corner of the wedge recedes diffusively,[6] that is $x, y \sim \sqrt{t}$. Because of the absence of any constant with the dimension of length in Eq. (9.40), the corresponding solution admits the self-similar form

$$X = x(t)/\sqrt{t}, \qquad Y(X) = y(x,t)/\sqrt{t}. \tag{9.42}$$

Notice that the increase of the magnetization is equal to twice the area under the curve $y(x,t)$. From (9.42), the area is proportional to t, so that the magnetization also grows linearly with time. To determine the evolution of the wedge, we substitute the ansatz (9.42) into the equation of motion (9.40) and find that the scaling function $Y(X)$ obeys

$$\frac{Y - XY'}{2} = \frac{Y''}{1 + (Y')^2}, \tag{9.43}$$

where the prime now indicates differentiation with respect to X. Equation (9.43) should be solved subject to the constraints $\lim_{X \to \infty} Y(X) = 0$ and $\lim_{X \to +0} Y(X) = \infty$, corresponding to the wedge geometry.

It is helpful to introduce the polar coordinates $(X, Y) = (r \cos\theta, r \sin\theta)$ to simplify Eq. (9.43) into the following equation for $r = r(\theta)$:

$$2r \frac{d^2 r}{d\theta^2} - \left(4 + r^2\right) \left(\frac{dr}{d\theta}\right)^2 = r^2 \left(2 + r^2\right). \tag{9.44}$$

Writing $dr/d\theta = R(r)$ reduces Eq. (9.44) to a first-order equation whose solution is $R^2 = r^4 \, e^{r^2/2} \, F(r, r_0)$, with

$$F(r, r_0) = \int_{r_0}^{r} \left(\frac{2}{\rho^3} + \frac{1}{\rho}\right) e^{-\rho^2/2} \, d\rho, \tag{9.45}$$

and r_0 is the scaled distance from the origin to the closest point on the interface. The locus of the interface is now determined from

$$\frac{dr}{d\theta} = \pm r^2 \, e^{r^2/4} \sqrt{F(r, r_0)}, \tag{9.46}$$

with the positive sign for $\theta \geq \pi/4$ and the negative sign for $\theta \leq \pi/4$ to give a symmetric solution about the diagonal $\theta < \pi/4$. Integrating Eq. (9.46) gives the explicit equation for $\theta = \theta(r)$:

$$\theta = \int_{r}^{\infty} \frac{e^{-\rho^2/4}}{\rho^2 \sqrt{F(\rho, r_0)}} \, d\rho \tag{9.47}$$

for $\theta \leq \pi/4$. For $\pi/4 < \theta < \pi/2$, we use $r(\theta) = r(\pi/2 - \theta)$ to ensure symmetry of the interface about the diagonal. The value of the unknown constant r_0 is obtained by imposing

---

[6] This feature applies both in the Ising–Glauber dynamics and in the continuum description.

the constraint that $\theta = \pi/4$ when $r = r_0$. This gives the criterion

$$\int_{r_0}^{\infty} \frac{e^{-r^2/4}}{r^2 \sqrt{F(r, r_0)}} \, dr = \frac{\pi}{4} \, , \tag{9.48}$$

whose numerical solution is $r_0 \doteq 1.0445$. Equation (9.47), with F given by (9.45), provides an explicit representation of $\theta(r)$ on the interface in terms of the (scaled) distance $r \in [r_0, \infty)$ from the origin. For $r \to \infty$, the interface equation simplifies considerably. In this limit, Eqs (9.45) and (9.47) give $\theta \to A r^{-3} e^{-r^2/4}$ with $A = 2/\sqrt{F(\infty, r_0)}$. In Cartesian coordinates, this relation between θ and r leads to is $Y \to A X^{-2} e^{-X^2/4}$.

9.4 Final states

In analogy with our earlier discussion in Sections 8.5 and 8.6 on the evolution of the zero-temperature Ising–Glauber model, there are some fundamental questions about coarsening in the long-time limit: (i) What is the final state of a system that evolves by zero-temperature TDGL dynamics? (ii) What are the low-energy excitations? There are similarities but also some important and subtle differences between the TDGL and Ising–Glauber dynamics that we now highlight.

In one dimension, we have seen (page 285) that the elementary TDGL excitations are kinks and antikinks; these are the analogs of domain walls in the Ising–Glauber model. In the Ising–Glauber system, domain walls diffuse freely and annihilate when they meet. Consequently, the time to reach the final ground state scales as the square of the system length. On the other hand, the interaction between kinks and antikinks in the TDGL equation is exponentially small when their separation exceeds their widths (this feature will be discussed in Section 9.7). Consequently, the dynamics is dominated by the pair of excitations with the smallest separation. This extremal dynamics picture, which will be treated in Section 9.7, leads to a slow evolution in which the time to reach the ground state by TDGL dynamics scales exponentially – rather than as a power law – in the size of the system.
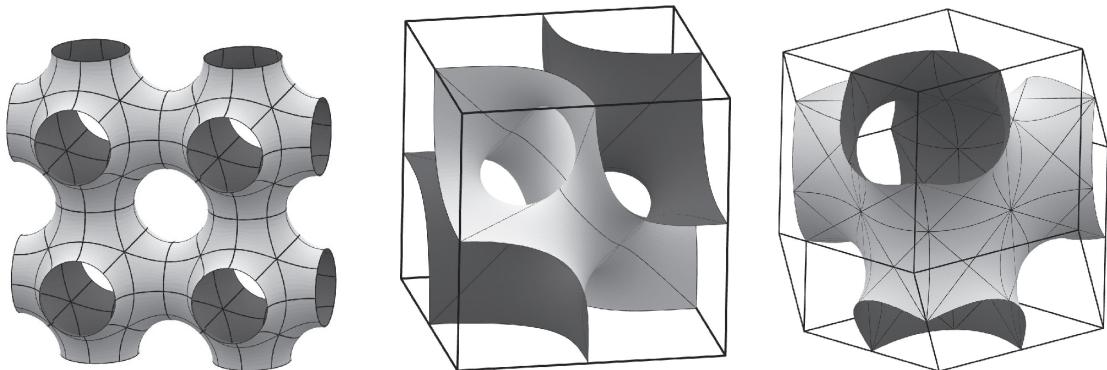
When $d \geq 2$, the outcomes of zero-temperature TDGL and Ising–Glauber dynamics are initially similar. Both dynamics appear to give similar patterns for coarsening that evolve in the same power-law fashion. Moreover, when the two systems are prepared in the same initial state, they both appear to reach the same final state (either the ground state or a stripe state) with the *same* probabilities for these two possible outcomes. Since the TDGL equation was constructed to be the natural macroscopic counterpart of microscopic Glauber dynamics, this correspondence might be expected. However, the zero-temperature TDGL equation is deterministic while Glauber dynamics remains stochastic even at zero temperature. Thus the nature of the connection between the two dynamics merits a closer look.

There are two subtleties involved. First, because TDGL evolution is deterministic, the final state is uniquely determined by the initial condition. On the other hand, starting with a given initial state and letting the system evolve by zero-temperature Glauber dynamics, different realizations of the dynamics can lead to different final states. Hence to determine

the probability to reach a stripe state, we must average over many initial conditions that all have zero magnetization. However, if we start with a "sufficiently random" initial condition, we may merely average over many realizations of the dynamics.[7]

A second subtlety is the huge separation of time scales in TDGL dynamics between the coarsening regime and the final state. If a stripe state is reached in TDGL dynamics, this state is attained in a time that scales as L^2, where L is the linear dimension of the system. In this stripe state, the exponential interaction between adjacent walls causes the system to eventually condense into the ground state – the true final state of the dynamics. However, this final state is reached in a time that scales as e^L rather than as a power law in L. For a typical macroscopic sample, say $L = 10^8$, this lifetime exceeds any time scale in the universe so this extraordinarily slow process can be disregarded.

In three dimensions, the final[8] states are reminiscent of those that arise in Ising–Glauber dynamics (see Fig. 8.10), although the similarity is a bit less precise than in two dimensions. The Allen–Cahn equation tells us that the final states consist of static interfaces with zero mean curvature. Such states are called minimal surfaces because they have the smallest possible area for a surface spanning the boundary. Soap films are real-life examples of minimal surfaces. For TDGL dynamics on a periodic domain, the minimal surface extends into the entire space $\mathbb{R}^3$, leading to the so-called triply periodic minimal surfaces.[9] Some simple examples of minimal surfaces are shown in Fig. 9.4. Even though the interface has zero mean curvature, there again is a residual exponentially small interaction between interfaces that should ultimately lead the system to reach the ground state. Another difference

Fig. 9.4. Examples of triply periodic minimal surfaces in three dimensions. Figures kindly provided by K. Brakke.

[7] The term "sufficiently random" is hard to specify. An example of a "bad" initial state is one that consists of two straight stripes; such a state is static in Glauber dynamics. On the other hand, the antiferromagnetic initial state – a completely deterministic state that has the highest energy – is probably a good starting point for Glauber dynamics.

[8] Again disregarding processes that occur on time scales of the order of e^L.

[9] The first five examples of triply periodic minimal surfaces were found by Schwarz and his students in the second half of nineteenth century; many examples have been found just since 1970.

with the Ising–Glauber model is that the TDGL system does not perpetually wander over an iso-energetic set of metastable states as in the discrete-spin model.

9.5 Defects

The domain wall discussed in Section 9.3 is the simplest *topological* defect. To appreciate its importance let us consider all of the possible solutions of the TDGL equation with *finite* free energy:

$$F[m] = \int dx \left[\frac{1}{2} \left(\frac{\partial m}{\partial x} \right)^2 + V(m) \right]. \tag{9.49}$$

At any time the magnetization must tend to a zero of the potential at spatial infinity; otherwise the free energy (9.49) would be infinite. For a one-dimensional system with two degenerate potential minima, say at $m = \pm 1$, the space of all solutions divides into four sectors that are characterized by the values of $m(x = \pm\infty)$. For instance, the kink sectors correspond to $m(x = -\infty, t) = \mp 1$ and $m(x = \infty, t) = \pm 1$. These sectors are topologically disconnected, as a path from a solution in one sector to a solution in a different sector would necessarily go through an infinite free energy[10] which is impossible – the free energy can only decrease when the evolution is described by the TDGL equation. Within each sector, the magnetization evolves so that the free energy decreases and eventually reaches a minimum. There are just four such minima – the stationary kink in the $[-1, 1]$ sector, the stationary antikink in the $[1, -1]$ sector, and one of the two ground states in each of the two remaining sectors. Thus domain walls (and ground states) have a special role as the final states of TDGL evolution.

Virial theorem

In higher dimensions, domain walls no longer have finite free energy, but they also do not lead to new physics as they are essentially one-dimensional. Are there stationary solutions with a *finite* free energy in higher dimensions? Strikingly, the answer is negative – when the order parameter is scalar, ground states are the only stationary solutions. This assertion is known as a virial (or Derrick's) theorem.

To prove this virial theorem suppose that $m = m_*(\mathbf{x})$ is a non-trivial stationary solution of the TDGL equation,

$$\nabla^2 m = V'(m), \tag{9.50}$$

that originates from minimizing the free energy functional, $\delta F = 0$; here ∇^2 is the d-dimensional Laplacian. It is useful to split the free energy functional into "kinetic" and

---

[10] For instance, if $m(x = \infty, t)$ changes continuously from $+1$ to -1, the free energy will be infinite whenever $m(x = \infty, t) \neq \pm 1$.

"potential" components: $F[m] = K[m] + P[m]$, with

$$K[m] = \int d\mathbf{x}\, \frac{1}{2}\, (\nabla m)^2 \geq 0, \qquad P[m] = \int d\mathbf{x}\, V(m) \geq 0.$$

Let us make a scale transformation of the stationary solution:

$$m_*(\mathbf{x}) \to m_a(\mathbf{x}) = m_*(a\mathbf{x}).$$

Under this transformation

$$K[m_a] = a^{2-d} K[m_*], \quad P[m_a] = a^{-d} P[m_*], \tag{9.51}$$

and therefore

$$F[m_a] = a^{2-d} K[m_*] + a^{-d} P[m_*]. \tag{9.52}$$

Since $m_1 \equiv m_*$ is the extremum of $F[m]$, it is also the extremum on a one-parameter family of magnetizations $m_a(\mathbf{x})$. Therefore

$$\left.\frac{dF[m_a]}{da}\right|_{a=1} = 0\,, \tag{9.53}$$

which in conjunction with (9.52) yields

$$(2 - d)K[m_*] = dP[m_*]\,. \tag{9.54}$$

Equation (9.54) together with inequalities $K[m_*] \geq 0$ and $P[m_*] \geq 0$ impose stringent constraints on the existence of stationary solutions:

1. *For $d > 2$.* Equation (9.54) is satisfied only when $K[m_*] = P[m_*] = 0$. The condition $= P[m_*] = 0$ shows that m_* must be constant and equal to one of the zeros of $V[m]$; i.e. the solution is the trivial ground state.
2. *For $d = 2$.* Equation (9.54) yields $P[m_*] = 0$. This condition is still sufficient (when the minima of $V[m]$ are discrete) to exclude non-trivial stationary solutions.
3. *For $d = 1$.* Equation (9.54) is satisfied when $K[m_*] = P[m_*]$. A similar relation arises in classical mechanics, where it is called the virial theorem. This connection explains the use of the "virial" terminology.

The absence of stationary solutions in $d > 2$ is also evident from (9.52). Indeed, this equation shows that $m_a(\mathbf{x})$ with $a > 1$ has a smaller energy than $m_* \equiv m_1(\mathbf{x})$. Thus the characteristic size of $m_a(\mathbf{x})$ is a times smaller than the size of $m_*(\mathbf{x})$. Therefore it is energetically favorable for the configuration $m_*(\mathbf{x})$ to shrink to a point, e.g. to become a ground state.

The virial theorem tells us that systems with a scalar order parameter admit non-trivial stationary solutions only in one dimension; it says nothing about time-dependent solutions. The latter do exist in certain models. Here we are interested only in time-independent solutions and therefore the only way to circumvent the virial theorem is to consider models with more complicated order parameters.

Vector order parameter

The next level of sophistication is a vector order parameter. This generalization is straightforward and all previous results continue to hold except for the replacement of the scalar order parameter by the vector $\mathbf{m} = (m_1, \ldots, m_n)$. For example, Eq. (9.54) becomes

$$(2 - d)K[\mathbf{m}_*] = dP[\mathbf{m}_*].\tag{9.55}$$

At first sight, nothing has changed compared to a scalar order parameter; e.g. non-trivial stationary solutions do not exist when $d > 2$. In two spatial dimensions, however, the situation is more interesting. Stationary solutions for the zero-potential system, $V[\mathbf{m}] \equiv 0$, satisfy the Laplace equation $\nabla^2 \mathbf{m} = 0$. Since (non-singular) solutions of the Laplace equation in infinite space are constants, the model with no potential does not possess non-trivial stationary solutions.

However, when we augment the potential-free model with the physical constraint

$$\mathbf{m}(\mathbf{x}) \cdot \mathbf{m}(\mathbf{x}) = \sum_{j=1}^{n}[m_j(\mathbf{x})]^2 = 1,\tag{9.56}$$

then non-trivial stationary states can arise. This constraint provides a more faithful macroscopic description of the (classical) Heisenberg ferromagnet that consists of fixed-length spins[11] than models with non-zero potential. Naively, we might anticipate that the governing equation is again $\nabla^2 \mathbf{m} = 0$, as follows from (9.50) and $V = 0$. However, this equation is erroneous: for systems with constraints, the proper quantity to minimize is the initial function plus the constraint, which is multiplied by a judiciously chosen Lagrange multiplier. In the present case, this procedure leads to the functional

$$F[\mathbf{m}] = \frac{1}{2} \int d\mathbf{x} \left\{ (\nabla \mathbf{m})^2 + \lambda[\mathbf{m}^2 - 1] \right\},$$

with an as yet unknown Lagrange multiplier $\lambda(\mathbf{x})$. Applying the Euler–Lagrange prescription $\delta F / \delta \mathbf{m} = 0$ to the above functional we obtain

$$\nabla^2 \mathbf{m} = \lambda \mathbf{m}.\tag{9.57}$$

Multiplying this equation by $\mathbf{m}$ and using the constraint (9.56) we express the Lagrange multiplier λ via the magnetization: $\lambda(\mathbf{x}) = \mathbf{m} \cdot \nabla^2 \mathbf{m}$. Therefore the governing equation (9.57) becomes

$$\nabla^2 \mathbf{m} = (\mathbf{m} \cdot \nabla^2 \mathbf{m})\, \mathbf{m}.\tag{9.58}$$

Unlike the linear Laplace equation $\nabla^2 \mathbf{m} = 0$, which would have emerged if there was no constraint, Eq. (9.58) is an elegant, but challenging, nonlinear partial differential equation.

---

[11] The model with zero potential and magnetization lying on the unit sphere in n dimensions is called the (nonlinear) $\mathcal{O}(n)$ model. The $\mathcal{O}(2)$ model is also known as the XY model and the $\mathcal{O}(3)$ model is often called the Heisenberg model.

Let's try to figure out when (9.58) has solutions without explicitly solving it. Owing to constraint (9.56) we can think of $\mathbf{m}(\mathbf{x})$ as a mapping from the two-dimensional physical space in which the model is defined onto a unit sphere $\mathbb{S}^{n-1}$ in the n-dimensional internal spin space:

$$\mathbf{m} : \mathbb{R}^2 \to \mathbb{S}^{n-1}. \tag{9.59}$$

Note that at spatial infinity the magnetization must approach some constant unit vector $\mathbf{m}_\infty$ in the internal spin space:

$$\lim_{|\mathbf{x}| \to \infty} \mathbf{m}(\mathbf{x}) = \mathbf{m}_\infty . \tag{9.60}$$

If there was no such boundary condition, the magnetization would depend on the polar angle θ, even when $r \to \infty$. This fact would imply that the angular component of the gradient $r^{-1} \partial \mathbf{m} / \partial \theta \sim r^{-1}$. Consequently, the free energy,

$$\frac{1}{2} \int r \, dr \, d\theta \, (\nabla \mathbf{m})^2 ,$$

would diverge (logarithmically).

Since $\mathbf{m}(\mathbf{x})$ must tend to the same value $\mathbf{m}_\infty$ for all points at spatial infinity, we can identify the two-dimensional plane $\mathbb{R}^2$ with the sphere $\mathbb{S}^2$. Consequently, (9.59) can be viewed as the mapping

$$\mathbf{m} : \mathbb{S}^2 \to \mathbb{S}^{n-1} \tag{9.61}$$

from the two-dimensional sphere onto the $(n-1)$-dimensional sphere.

Remarkably, such mappings form sectors; mappings from the same sector can be deformed into each other, while mappings from different sectors cannot. These sectors are called *homotopy groups* and are denoted by $\pi_2(\mathbb{S}^{n-1})$. The following groups are known:

1. If $n \geq 4$, there is only a single sector that contains the constant mapping. In this case, the homotopy groups are trivial: $\pi_2(\mathbb{S}^k) = 0$ when $k \geq 3$. This fact does not formally prove that there are only trivial stationary solutions for the order parameter, but at least makes this feature plausible.
2. When $n = 3$, i.e. for the Heisenberg model, there are infinitely many sectors. This feature can be expressed compactly as $\pi_2(\mathbb{S}^2) = \mathbb{Z}$, which says that sectors can be labeled by integers (and hints that homotopy classes can be added and form a group that is isomorphic to the group of integers $\mathbb{Z}$). The integer "topological" charge of the mapping (9.61) is given by the integral[12]

$$q = \frac{1}{8\pi} \int d\mathbf{x} \, [\epsilon_{ij} \mathbf{m} \cdot (\partial_i \mathbf{m} \times \partial_j \mathbf{m})] . \tag{9.62}$$

---

[12] It is not obvious that the right-hand side is an integer; the property that for *any* mapping $\mathbf{m} : \mathbb{S}^2 \to \mathbb{S}^2$ the integral in (9.62) divided by 8π is an integer is a typical, striking result of algebraic topology.

Here the indices label spatial coordinates: $d\mathbf{x} = dx_1 dx_2$, $\partial_i \mathbf{m} = \partial \mathbf{m}/\partial x_i$, etc.; repeated indices should be summed; and ϵ_{ij} denotes the antisymmetric tensor: $\epsilon_{12} = -\epsilon_{21} = 1$, $\epsilon_{11} = \epsilon_{22} = 0$.

3. Finally, $\pi_2(\mathbb{S}^1) = 0$ implies that for the *XY* model there are no non-trivial stationary solutions for an order parameter that has a finite free energy.

Stationary solutions of Heisenberg model

The above arguments suggest that non-trivial solutions with finite free energy, and in particular stationary solutions, exist in the Heisenberg model in two dimensions. Recall that the free energy of the Heisenberg model is

$$F[\mathbf{m}] = \frac{1}{2} \int d\mathbf{x}\, (\partial_i \mathbf{m}) \cdot (\partial_i \mathbf{m}), \tag{9.63}$$

with the order parameter equal to the three-dimensional unit vector $\mathbf{m} = (m_1, m_2, m_3)$, i.e. with $\mathbf{m} \cdot \mathbf{m} = 1$. Each solution with a finite free energy belongs to one of the sectors; the total number of sectors is infinite and an evolving solution is confined to a single sector. It is difficult to find a solution within any sector because even the simplest stationary solutions must satisfy the formidable looking equation (9.58). The only exception is the trivial ground-state solution $\mathbf{m}(\mathbf{x}) \equiv \mathbf{m}_\infty$,[13] where the corresponding free energy vanishes, $F = 0$.

Some non-trivial solutions, however, have been constructed by using a pair of clever tricks.

First trick

We begin with the obvious inequality

$$\int d\mathbf{x}\, [(\partial_i \mathbf{m} \pm \epsilon_{ij} \mathbf{m} \times \partial_j \mathbf{m}) \cdot (\partial_i \mathbf{m} \pm \epsilon_{ik} \mathbf{m} \times \partial_k \mathbf{m})] \geq 0.$$

Expanding this inequality we obtain

$$\int d\mathbf{x}\, [(\partial_i \mathbf{m}) \cdot (\partial_i \mathbf{m}) + \epsilon_{ij} \epsilon_{ik} (\mathbf{m} \times \partial_j \mathbf{m}) \cdot (\mathbf{m} \times \partial_k \mathbf{m})]$$

$$\geq \pm 2 \int d\mathbf{x}\, [\epsilon_{ij} \mathbf{m} \cdot (\partial_i \mathbf{m} \times \partial_j \mathbf{m})]. \tag{9.64}$$

The terms on the left-hand side are equal to each other:

$$\epsilon_{ij} \epsilon_{ik} (\mathbf{m} \times \partial_j \mathbf{m}) \cdot (\mathbf{m} \times \partial_k \mathbf{m}) = \delta_{jk} [(\mathbf{m} \cdot \mathbf{m})(\partial_j \mathbf{m} \cdot \partial_k \mathbf{m}) + (\mathbf{m} \cdot \partial_j \mathbf{m})(\mathbf{m} \cdot \partial_k \mathbf{m})]$$

$$= (\partial_i \mathbf{m} \cdot \partial_i \mathbf{m}).$$

---

[13] There are infinitely many ground states, each of which is described by a unit vector $\mathbf{m}_\infty$. This indeterminacy is the manifestation of "spontaneous symmetry breaking" at the classical level.

(Here we use the identity $\epsilon_{ij}\epsilon_{ik} = \delta_{jk}$, the constraint $\mathbf{m} \cdot \mathbf{m} = 1$, and its derivative $\mathbf{m} \cdot \partial_j \mathbf{m} = 0$.) Therefore (9.64) becomes

$$\int d\mathbf{x} \, (\partial_i \mathbf{m}) \cdot (\partial_i \mathbf{m}) \geq \pm \int d\mathbf{x} \, [\epsilon_{ij} \mathbf{m} \cdot (\partial_i \mathbf{m} \times \partial_j \mathbf{m})] \,. \tag{9.65}$$

By recalling expression (9.62) for the topological charge and Eq. (9.63), we can rewrite (9.65) in the compact form

$$F \geq 4\pi |q| \,. \tag{9.66}$$

Hence within each q-sector the free energy can never be smaller than $4\pi |q|$. Recall that the stationary solutions are found by minimizing $F[\mathbf{m}]$ and the minimization can be done within each sector separately. Thus the minimal possible value of $F[\mathbf{m}]$ in the q-sector is $4\pi |q|$. The derivation of (9.65) shows that this minimal solution arises if and only if

$$\partial_i \mathbf{m} = \pm \epsilon_{ij} \mathbf{m} \times \partial_j \mathbf{m} \,. \tag{9.67}$$

This is a first-order partial differential equation and therefore represents a huge simplification in comparison to the original second-order partial differential equation (9.58). It is, in principle, possible that (9.58) has stationary solutions that correspond to local minima of $F[\mathbf{m}]$ that exceed the global minimum $4\pi |q|$. We focus, however, on the solutions that arise from the much simpler first-order equation (9.67).

Second trick

The goal now is to reduce the nonlinear equation (9.67) to a linear one. The idea is to use a stereographic projection of the sphere $\mathbf{m} \cdot \mathbf{m} = 1$ on the plane. We choose the plane that is parallel to (m_1, m_2) and contains the south pole. The coordinates (y_1, y_2) of this plane are related to $\mathbf{m}$ via

$$y_1 = \frac{2m_1}{1 - m_3}, \quad y_2 = \frac{2m_2}{1 - m_3} \,. \tag{9.68}$$

Now using (9.67) we calculate $\partial_1 y_1$ and $\partial_2 y_2$ and find that they coincide (or anti-coincide), and similarly for the two other derivatives. In summary,

$$\frac{\partial y_1}{\partial x_1} = \pm \frac{\partial y_2}{\partial x_2}, \qquad \frac{\partial y_1}{\partial x_2} = \mp \frac{\partial y_2}{\partial x_1} \,. \tag{9.69}$$

These are not merely linear equations, these are the Cauchy–Riemann equations! They tell us that the complex variable $\zeta = y_1 + iy_2$ is an analytic function of $z^* = x_1 - ix_2$ for the upper signs in (9.69); for the lower signs, ζ is an analytic function of $z = x_1 + ix_2$.

Let's summarize our findings:

1. The goal was to find solutions of the nonlinear second-order partial differential equation (9.58) that represent stationary solutions of the Heisenberg model with finite free energy.

2. Using the first trick we arrive at the much simpler problem of finding solutions to the first-order equation (9.67). Such solutions satisfy the initial equation (9.58), although not all solutions of Eq. (9.58) are solutions of (9.67).

3. Using the second trick we find that, after the mapping (9.68), Eqs (9.67) reduce to the Cauchy–Riemann equations. Hence *any* analytical function gives a solution.

Vortices

We now give examples of stationary solutions. Because of their localized nature, it is sensible to call them vortices for any value of q. Such solutions are represented by analytic functions $\zeta = \zeta(z)$, with $\zeta = y_1 + iy_2$ and (y_1, y_2) given by Eq. (9.68). In terms of this function, the free energy (9.63) and the topological charge q are

$$F = \int d\mathbf{x} \, \frac{|d\zeta/dz|^2}{(1 + |\zeta|^2/4)^2} \qquad \text{with} \quad |q| = \frac{F}{4\pi}. \tag{9.70}$$

As an example, consider

$$\zeta = z^n$$

with n a positive integer. Substituting this into (9.70) we find $F = 4\pi n$ and $|q| = n$. If we choose an apparently more general analytic function $\zeta = [(z - z_0)/R]^n$, we get the same result for F. Here z_0 gives the location of the vortex; the independence of F on z_0 reflects translational invariance. Similarly, the parameter R is the size of the vortex and the independence of F on R reflects the invariance of the free energy under the scale transformation $\mathbf{x} \to R\mathbf{x}$.

The simplest vortex corresponding to $\zeta = 2z$ has a neat explicit representation in terms of the original variables

$$m_1 = \frac{2x_1}{|\mathbf{x}|^2 + 1}, \qquad m_2 = \frac{2x_2}{|\mathbf{x}|^2 + 1}, \qquad m_3 = \frac{|\mathbf{x}|^2 - 1}{|\mathbf{x}|^2 + 1}. \tag{9.71}$$
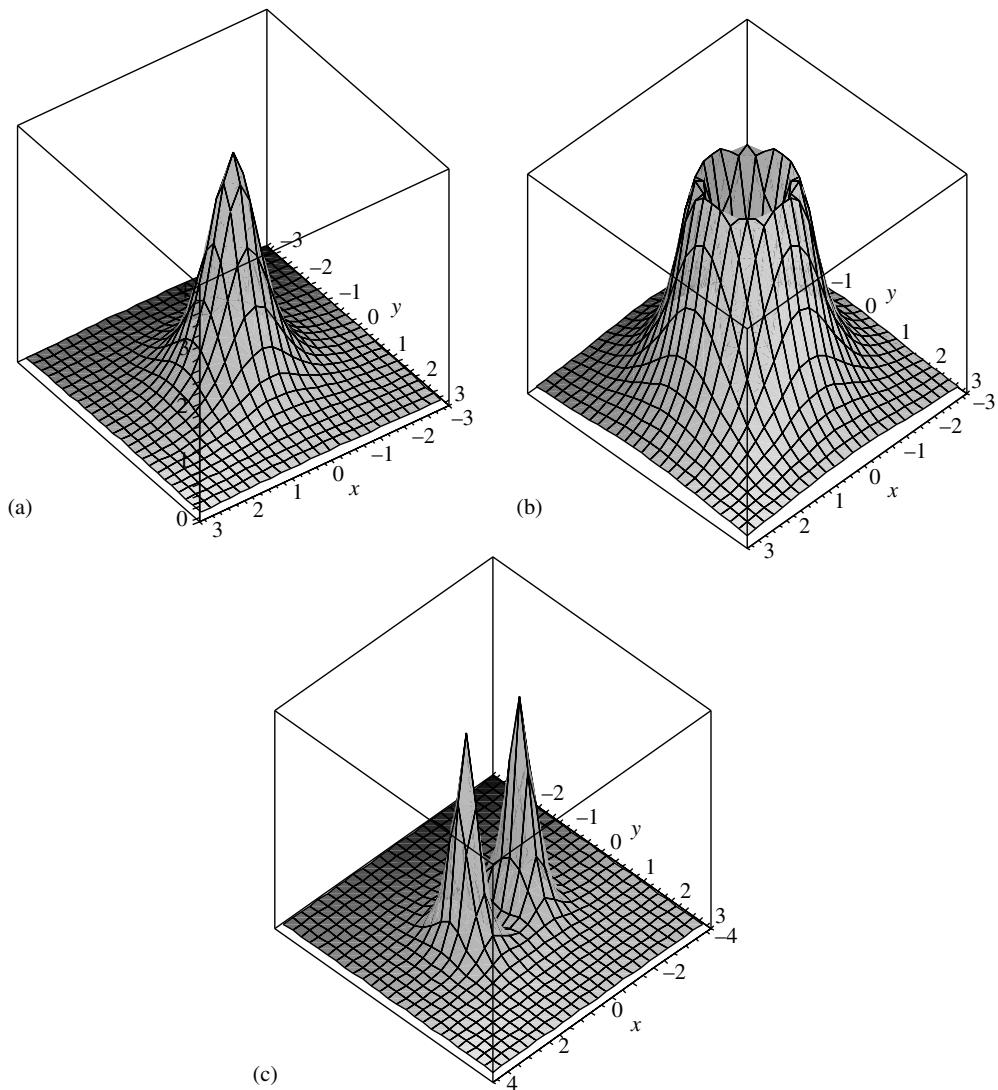
Far away from the center of this vortex (the center is set to be at the origin), the magnetization approaches $\mathbf{m}_\infty = (0, 0, -1)$. The topological charge of the vortex (9.71) is $q = -1$; a vortex with charge $q = 1$ is obtained, for example, by reversing the sign of m_3 in (9.71). This gives

$$m_1 = \frac{2x_1 R}{|\mathbf{x}|^2 + R^2}, \qquad m_2 = \frac{2x_2 R}{|\mathbf{x}|^2 + R^2}, \qquad m_3 = \frac{R^2 - |\mathbf{x}|^2}{|\mathbf{x}|^2 + R^2}, \tag{9.72}$$

where we additionally restored R, the vortex size, which can be arbitrary. The most general solution for the vortex with charge $q = 1$ coincides with (9.72) up to translation and rotation.

It is not easy to visualize $\mathbf{m}(\mathbf{x})$. Instead of plotting this vector function, let us plot the scalar energy density $\mathcal{F}$, defined by $F = \int d\mathbf{x} \, \mathcal{F}$ (Fig. 9.5). According to (9.70),

$$\mathcal{F} = \frac{|d\zeta/dz|^2}{(1 + |\zeta|^2/4)^2}. \tag{9.73a}$$

(a)

(b)

(c)

The energy densities (a) $\mathcal{F}_1$, (b) $\mathcal{F}_2$, and (c) $\mathcal{F}_{11}$ (with $a = 1$) in two dimensions as a function of (x, y). These densities are associated with vortices of topological charge $q = -1, -2$, and -2, respectively.

For $\zeta = 2z$, the energy density becomes

$$\mathcal{F}_1 = \frac{4}{(1 + |z|^2)^2} = \frac{4}{(1 + x_1^2 + x_2^2)^2} \,. \tag{9.73b}$$

The subscript indicates that this energy density corresponds to a vortex of topological charge $q = -1$. Similarly taking $\zeta = 2z^2$ gives a vortex with charge $q = -2$, whose energy

density is

$$\mathcal{F}_2 = \frac{16|z|^2}{(1 + |z|^4)^2} = \frac{16(x_1^2 + x_2^2)}{[1 + (x_1^2 + x_2^2)^2]^2} . \tag{9.73c}$$

Finally, taking $\zeta = 2(z - a)(z + a)$ gives a (nonlinear) superposition of $q = -1$ vortices to give an excitation with topological charge $q = -2$ and energy density

$$\mathcal{F}_{11} = \frac{16(x_1^2 + x_2^2)}{(1 + [(x_1 - a)^2 + x_2^2][(x_1 + a)^2 + x_2^2])^2} . \tag{9.73d}$$

When $a \gg 1$, the constituent vortices are approximately located at $(x_1, x_2) = (\pm a, 0)$.

Significance of vortices

To appreciate the significance of vortices let us first recall some basic features of domain walls in one dimension. Because a domain wall has a finite free energy, there will be a finite density of these walls[14] at any positive temperature. The existence of a finite density of domain walls destroys long-range order at any positive temperature in one dimension. Localized vortices with finite free energy that arise in the classical Heisenberg model in two dimensions play a similar role: their mere existence shows that their density is of the order of $e^{-F/T}$ for any positive temperature. Thus vortices allow us to establish a crucial feature of the equilibrium behavior of the Heisenberg model in two dimensions – long-range order is impossible at any positive temperature.

Vortices are much more diverse in character than domain walls. Indeed, there are infinitely many types of vortices (each with a topological charge that is an arbitrary integer); moreover, vortices with the same charge can differ in size. In contrast, there are just two types of domain walls[15] and their sizes are all equal. The presence of many types of vortices distinguishes the Heisenberg model from the Ising model. However, at low temperatures almost all vortices will have charge $q = \pm 1$. In this respect, the situation is similar to the one-dimensional Ising model which also possesses two types of domain walls – kinks and antikinks.

Domain walls and vortices deviate in several other important aspects. For instance, in one dimension kinks alternate with antikinks; there is no such topological restriction on the location of vortices in two dimension. It is also much more difficult to describe the interaction between vortices than between domain walls. Since the governing equations are nonlinear, we can talk about well-defined vortices only when their separation L is large, that is, when L greatly exceeds their sizes: $L \gg R_1 + R_2$. Such widely separated vortices would approximately maintain their identity, apart from a small distortion. The two vortices will attract or repel each other. Vortices with opposite topological charge can probably annihilate; the details of this type of interaction are unknown. On the other hand, a kink and antikink attract each other, with an interaction that scales exponentially with their separation (see Section 9.7).

---

[14] Of the order of $e^{-1/T}$ if we set the energy of a domain wall to one.

[15] This statement applies to potentials with two degenerate minima. For potentials with m degenerate minima, there are $2(m - 1)$ types of domain walls.
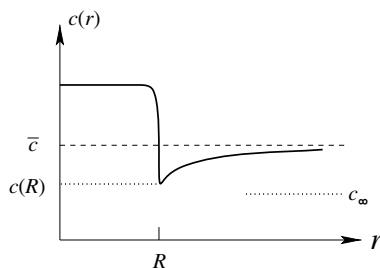
9.6 Conservative dynamics

The influence of a conservation law severely limits the way in which an interface between two domains can move. While an individual interface can move to reduce its local curvature and energy, such an evolution has to be accompanied by a global rearrangement of interfaces to ensure that the order parameter is conserved. Thus, even the evolution of a single droplet is non-trivial in the conservative case. We therefore begin our discussion with a single droplet and then continue to the evolution of a dilute population of droplets that are immersed in a majority phase. In the dilute limit, droplets are widely separated and therefore almost spherical. By treating droplets as spherical throughout the evolution, we can avoid the complexities associated with random interfaces. Nevertheless, this limit is still quite subtle – droplets smaller than the average size shrink while larger droplets grow, and the average droplet size must be determined self-consistently.

Evolution of a single droplet

Consider the evolution of a single droplet of one phase that is immersed in a background of the opposite phase. Already at the problem statement level, there is an inconsistency: if the two phases were perfectly separated, then the droplet could not evolve because any change in the droplet size would violate order-parameter conservation. Instead, we should think of a single liquid droplet together with its vapor in a closed container. At the surface of the droplet both the evaporation as well as the condensation of molecules of vapor back into the droplet can occur. Through the combination of these two processes, the droplet can evolve while still conserving the order parameter.

   The condensation rate is given by the flux of vapor molecules to the droplet which, in turn, is determined by the vapor concentration in the gas. In principle, the vapor concentration exterior to the droplet obeys the diffusion equation. However, because the droplet radius changes slowly with time, we apply the quasi-static approximation (see Section 2.7) in which we ignore the explicit time dependence and deal with the simpler Laplace equation to determine the external vapor concentration. The concentration of the minority phase as a function of distance from the center of a droplet of radius R should have the dependence sketched in Fig. 9.6. The average concentration $\bar{c}$ of the minority phase in the entire system

Fig. 9.6.   Schematic (not to scale) dependence of the minority-phase concentration as a function of the radial coordinate.

must be larger than the supersaturation value c_∞. That is, there must be supersaturation so that a droplet can form in the first place. (For $\bar{c} < c_\infty$, the minority phase remains as a homogeneous vapor in the gas.) Inside the droplet, the concentration is much greater than $\bar{c}$ by the very definition of a droplet. Outside the droplet, the vapor concentration obeys Laplace's equation with the boundary conditions $c(r \to \infty) \to \bar{c}$ and $c(R)$ determined by the Gibbs–Thompson relation (see below). In three dimensions, $c(r)$ therefore varies as

$$c(r) = \bar{c} - [\bar{c} - c(R)]\,\frac{R}{r}\,. \tag{9.74}$$

The Gibbs–Thompson relation relates $c(R)$ to c_∞ by the following physical picture: For a bulk liquid in equilibrium with a saturated vapor, the vapor concentration at the interface equals c_∞ by definition. For a small droplet, the vapor concentration at the interface is larger than that of a bulk liquid because a larger fraction of molecules are at the interface. Thus $c(R)$ should exceed c_∞ by an amount that vanishes as the droplet radius goes to infinity. This relationship is encapsulated by the Gibbs–Thompson relation $c(R) = c_\infty\,(1 + v/R)$, where v is known as the capillary length that is related to the surface tension and the temperature.

Assembling these elements, the volume V of an isolated droplet changes with rate

$$\frac{dV}{dt} = 4\pi R^2\,\frac{dR}{dt} = 4\pi R^2\,D\frac{\partial c}{\partial r}\bigg|_{r=R}\,, \tag{9.75}$$

from which $dR/dt = D[\bar{c} - c(R)]/R$. We now define $\Delta = \bar{c} - c_\infty$ as the degree of supersaturation, and use the Gibbs–Thompson relation to eliminate $c(R)$ and give

$$\frac{dR}{dt} = \frac{D}{R}\left(\Delta - c_\infty\frac{v}{R}\right) \equiv \frac{\alpha}{R^2}\left(\frac{R}{R_c} - 1\right). \tag{9.76}$$

From the latter form, it is clear that for $R > R_c = c_\infty v/\Delta$ the droplet grows, while for $R < R_c$ the droplet shrinks. As a particularly simple example, consider the limiting case $\bar{c} = c_\infty$. Then $R_c = \infty$ and Eq. (9.76) reduces to $dR/dt = -\alpha R^{-2}$. Thus the droplet shrinks, $R^3 = R_0^3 - 3\alpha t$. This behavior is the building block of $t^{1/3}$ coarsening that is characteristic of conserved order-parameter dynamics.

Lifshitz–Slyozov–Wagner coarsening

The foregoing discussion provides the basis for understanding the coarsening of a dilute population of droplets in a supersaturated background. This phenomenon (analyzed by Lifshitz, Slyozov, and Wagner) is known as LSW coarsening. The dilute limit allows us to make the approximation that droplets are non-interacting, so that the concentration field around each droplet is the same as that of an isolated droplet. The basic feature of this coarsening is already contained in Eq. (9.76): droplets whose radii exceed R_c grow and those with smaller radii shrink. In a population of heterogeneous droplets, the value of R_c has to be determined self-consistently, and we show that this leads to $R_c \sim t^{1/3}$.

Let $f(R, t)$, be the concentration of droplets of radius R at time t. This concentration evolves by the continuity equation

$$\frac{\partial f}{\partial t} + \frac{\partial J}{\partial R} = 0, \tag{9.77}$$

where the flux $J = f(R, t)\, dR/dt$ is just the difference between the increase and decrease of droplets of radius R due to their evolution. We wish to solve this equation of motion with dR/dt given by Eq. (9.76) and subject to the constraint that the total mass of the minority phase is conserved. This constraint may be written as

$$\bar{c} - c_\infty + \frac{4\pi}{3} \int_0^\infty R^3 f(R, t)\, dR = \text{const.} \tag{9.78}$$

In the minority limit, the volume fraction of the minority phase that exists as freely diffusing monomers is vanishingly small. Thus the conservation law reduces to the condition that the total volume of the droplets is fixed. With this proviso, Eqs (9.76)–(9.78) constitute the governing equations of conserved order-parameter coarsening in the minority limit.

To solve these equations of motion for coarsening, we again apply scaling. The natural scaling ansatz for this system is

$$f(R, t) = \frac{1}{R_c^4}\, \phi(z), \qquad z = \frac{R}{R_c}. \tag{9.79}$$

Here the prefactor is determined by the conservation of the total mass of the minority phase, namely, $\int R^3 f(R, t)\, dR = \text{const.}$, which in conjunction with (9.79) gives the relation $\int z^3 \phi(z)\, dz = \text{const.}$, which is manifestly time-independent. Substituting the scaling ansatz (9.79) into (9.77), the first term in this equation becomes

$$\frac{\partial f}{\partial t} = -\frac{4\dot{R}_c}{R_c^5}\, \phi - \frac{1}{R_c^4}\, \phi'\, \frac{R\dot{R}_c}{R_c^2} = -\frac{\dot{R}_c}{R_c^5}(4\phi + z\phi'),$$

where the prime denotes differentiation with respect to the scaled variable $z \equiv R/R_c$. Similarly, the second term in (9.77) becomes

$$\frac{\partial}{\partial R}\left[\frac{\alpha}{R^2}\left(\frac{R}{R_c} - 1\right)\frac{1}{R_c^4}\, \phi \right] = \frac{\alpha}{R_c^7}\left[\left(\frac{1}{z} - \frac{1}{z^2}\right)\phi' + \left(\frac{2}{z^3} - \frac{1}{z^2}\right)\phi \right].$$

With these preliminaries, we convert the partial differential equation (9.77) into the ordinary differential equation

$$R_c^2 \dot{R}_c = \alpha\, \frac{\left[\left(\dfrac{1}{z} - \dfrac{1}{z^2}\right)\phi' + \left(\dfrac{2}{z^3} - \dfrac{1}{z^2}\right)\phi \right]}{4\phi + z\phi'} \equiv \alpha\gamma, \tag{9.80}$$

where we have rearranged the terms to put all the time dependence on the left and all the z dependence on the right. Since both sides are functions of different variables they each must be constant. In elementary applications of scaling, such as the scaling solution to the

diffusion equation in Section 1.1, the value of the separation constant plays little role in the scaling solution. In the present case, the separation constant γ is crucial: depending on the value of γ, there are three different regimes of behavior, only one of which is physically meaningful.

Let us now examine the separated equations and determine the condition that gives the physical value of γ. For the time dependence we have

$$R_c^2 \dot{R}_c = \alpha\gamma \quad \Longrightarrow \quad R_c(t) = (3\alpha\gamma t)^{1/3} \ . \tag{9.81}$$

Thus coarsening under the constraint of a conserved order parameter leads to $t^{1/3}$ growth of the typical droplet radius. This slower-than-diffusive growth also justifies the quasi-static approximation that was used to determine the concentration outside a droplet. For the z dependence of the scaling function, the governing equation (9.80) becomes, after some rearrangement,

$$g\phi' - h\phi = 0, \quad \text{with} \quad g(z) = \frac{1}{z} - \frac{1}{z^2} - z\gamma, \quad h(z) = \frac{1}{z^2} - \frac{2}{z^3} + 4\gamma \ . \tag{9.82}$$

Thus the scaling function is formally given by

$$\ln\phi = \int^z \frac{h(y)}{g(y)}\, dy = \int^z \frac{2 - y - 4\gamma y^3}{1 - y + \gamma y^3} \frac{dy}{y} \ . \tag{9.83}$$

Thus far, our reasoning is standard: we've used scaling to separate the partial differential equation (9.77) into two ordinary differential equations. As often occurs, the time dependence follows easily. For LSW coarsening, the analysis of the z dependence is subtle because of the essential role played by the separation constant γ. A surprising consequence of Eq. (9.83) is that $\phi(z)$ must have a sharp cutoff at a value z_{max}, beyond which $\phi(z) = 0$. To establish this assertion, suppose the opposite is true. Then as $z \to \infty$, Eq. (9.83) would tell us that $\phi(z)$ should asymptotically vary as

$$\ln\phi \simeq -\int^z 4\frac{dy}{y} = -4\ln z \longrightarrow \phi \sim z^{-4} \ .$$
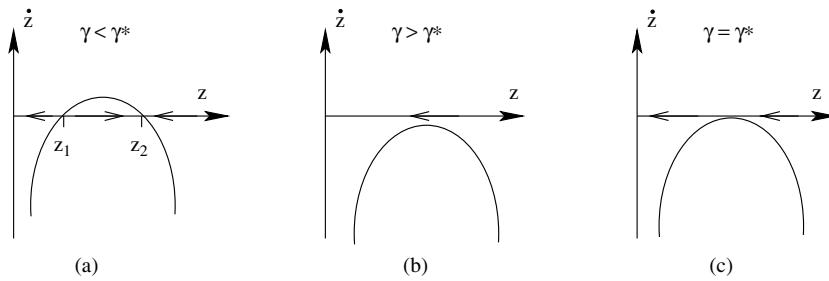
A power-law tail for ϕ is impossible, however, as this asymptotic decay would lead to a divergence in the total minority-phase mass:

$$\int R^3 \phi(R,t)\, dR \sim \int z^3 z^{-4}\, dz \longrightarrow \infty.$$

Thus we conclude that $\phi(z)$ has a sharp cutoff beyond some value z_{max}.

A second and more profound feature is that only one value of the separation constant γ is physically acceptable. To see why this is the case, let's re-examine the behavior of $\dot{R}$ in scaled units. Using Eqs (9.76) and (9.81), we find, after some simple algebra,

$$\dot{z} = \frac{1}{3\gamma t}\, g(z), \tag{9.84}$$

Sketch of $\dot{z} = g(z)/(3\gamma t)$ versus z for the three cases: (a) $\gamma < \gamma^*$, (b) $\gamma > \gamma^*$, and (c) $\gamma = \gamma^*$. The arrows on the z axis show the flow of z in Eq. (9.84).

with $g(z)$ defined in (9.82). Let's now examine the behavior of $g(z)$ for different values of γ. There are three distinct cases (see Fig. 9.7):

(a) $\gamma < \gamma^* = 4/27$. Here $z(t)$ flows exponentially quickly to the stable fixed point at z_2 for $z(0) > z_1$. That is, the radii of all droplets approach the common value $R_c z_2$, which diverges as $t \to \infty$. Such a distribution cannot satisfy mass conservation.

(b) $\gamma > \gamma*$. Here $z(t) \to 0$ exponentially quickly in time for any initial value of z. Thus all droplets shrink to zero and mass conservation is again violated.

(c) $\gamma = \gamma^*$. In this case $z = 3/2$ is a fixed point, but one that is approached as a power law in time. This subtle behavior is the mechanism that allows mass conservation to be satisfied. If the fixed point at $z = 3/2$ was reached exponentially quickly, then again all droplet radii would approach the common value of $3R_c/2$ and mass conservation would be violated. The slow decrease in z ensures the delicate balance between growth and shrinking of clusters in a mass-conserving way.

For the physical case of $\gamma = \gamma^*$, Eq. (9.83) for the scaling function can be factorized as

$$\ln \phi = \int^z \frac{2 - y - \frac{16}{27}y^3}{(\frac{2}{3}y - 1)^2(\frac{1}{3}y + 1)} \frac{dy}{y} \ .$$

Evaluating the latter integral by a partial fraction expansion gives the intricate form of the scaling function in three dimensions (Fig. 9.8):

$$\phi(z) = \begin{cases} Cz^2 \, (z+3)^{-7/3} \, (3-2z)^{-11/3} \, e^{-3/(3-2z)}, & z < 3/2, \\ 0, & z > 3/2, \end{cases} \qquad (9.85)$$

in which the amplitude C is fixed by the conservation law (9.78). The same qualitative phenomenology arises for LSW coarsening in arbitrary spatial dimensions (problems 9.15 and 9.16).
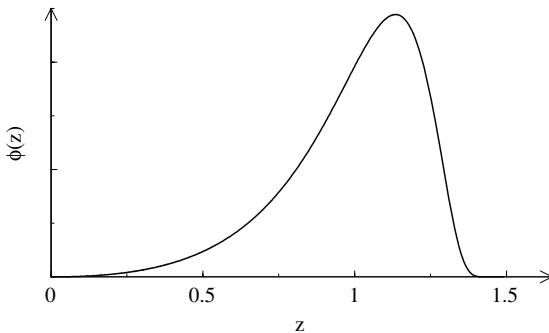
Fig. 9.8.   The scaling function $\phi(z)$ for LSW coarsening, Eq. (9.85).

9.7 Extremal dynamics

While power-law domain growth is a generic feature of zero-temperature coarsening in higher dimensions, one dimension is special because much slower logarithmic growth occurs. This difference is particularly evident in the case of non-conservative dynamics. In one dimension there is no local curvature to drive an interface; rather the interface moves by a net flux of order parameter across neighboring domains that is determined, in turn, by their lengths. In the long-time limit, this flux vanishes exponentially with the length of a domain and leads to *extremal dynamics*, in which the shortest domain in the system evolves, while all other domains (apart from the neighbors) hardly change. This extremal picture provides a natural way to determine the evolution of domains for both non-conservative and conservative dynamics. The net result of this dynamics is that domains coarsen logarithmically with time.

Shrinking of a single domain

To understand the origin of the slow domain growth in the framework of the one-dimensional TDGL equation, consider a single large domain of length $L = x_2 - x_1$ and magnetization close to $+1$ that is immersed in a sea with magnetization close to -1 (Fig. 9.9). The two interfaces of this bubble consist of a kink at $x = x_1(t)$ and an antikink at $x = x_2(t) = x_1 + L$.
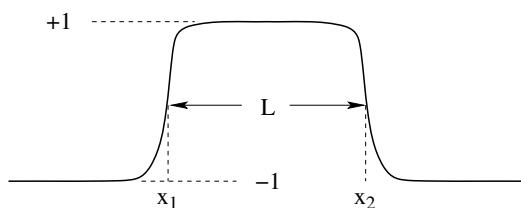
Fig. 9.9.   A single domain with magnetization close to $+1$ in a background where the magnetization is close to -1.

If L is much larger than the width of each interface, which is a good approximation at long times, the kink and antikink are nearly independent. Under this assumption, the spatial dependence of the magnetization can be written as

$$m(x, t) \simeq \tanh[x - x_1(t)] - \tanh[x - x_2(t)] - 1 \,, \tag{9.86}$$

where we have used the profile (9.23) for the kink and antikink.

Let's now estimate how the interfaces move for this kink–antikink pair. Substituting the profile (9.86) into the TDGL equation (9.2) we obtain

$$\frac{\dot{x}_1}{[\cosh(x - x_1)]^2} - \frac{\dot{x}_2}{[\cosh(x - x_2)]^2} = 6[\tanh(x - x_1) - \tanh(x - x_2)]$$
$$\times [1 - \tanh(x - x_1)][1 + \tanh(x - x_2)]. \tag{9.87}$$

Keeping only the lowest-order terms in $x - x_1$ and $x - x_2$, we find that asymptotically

$$\dot{x}_1 \simeq -\dot{x}_2 = 12 \, e^{-2(x_2 - x_1)}. \tag{9.88}$$

Consequently the domain length $L = x_2 - x_1$ shrinks according to $dL/dt \simeq -24 \, e^{-2L}$ whose solution is $L(t) = \frac{1}{2} \ln \left[e^{2L(0)} - 48t \right]$. The time for the domain to disappear is thus anomalously long: $\tau = \frac{1}{48} e^{2L(0)}$. This exponential length dependence of the shrinking time has profound consequences for the evolution of a heterogeneous domain array in one dimension. In the long-time limit, the shrinking time of the smallest domain is orders of magnitude shorter than that of the next smallest domain. As a result, domains merge with their neighbors in a size-ordered, and hence deterministic, way. This ordering leads to extremal dynamics and suggests the following efficient algorithm for TDGL domain evolution: (i) pick the smallest domain and merge it with its two nearest neighbors while keeping all other domains fixed; (ii) repeat *ad infinitum* (Fig. 9.10).
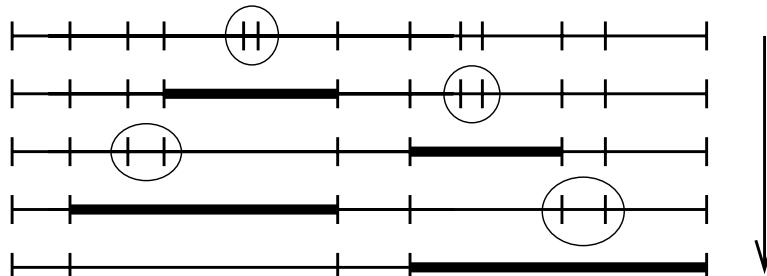
Fig. 9.10. Evolution in extremal dynamics. The shortest domain at any stage (circled) merges with its two neighbors, while all other domains remain static. The newly created domain at each step is highlighted.

The domain length distribution

We now study how the domain length distribution evolves by extremal dynamics.[16] More concretely, we determine $c(x, \ell)$, the density of domains of length x when the shortest domain has length ℓ. This minimal length plays the role of a time variable for extremal dynamics. We choose the initial domain lengths from a continuous distribution so that each domain has a distinct length. We also assume that the total initial magnetization vanishes, so that the average length of domains of positive and negative magnetization are equal. By definition, the total density of domains is $\rho = \int_\ell^\infty c(x, \ell)\, dx$, while the condition that the domains cover the entire line leads to $\int_\ell^\infty x\, c(x, \ell)\, dx = 1$.

Let's first show how the shortest domain plays the role of a time-like variable. When the shortest domain has length ℓ, suppose that the requisite number of mergings occur to eliminate all domains with lengths in the range $(\ell, \ell + \Delta\ell)$. Consequently, the length of the shortest domain increases from ℓ to $\ell + \Delta\ell$. The density of domains in this length range is $c(\ell, \ell)\Delta\ell$. Since there is a net loss of two domains in each merger, $\rho(\ell + \Delta\ell) = \rho(\ell) - 2c(\ell, \ell)\Delta\ell$, and the domain density obeys

$$\frac{d\rho}{d\ell} = -2c(\ell, \ell)\,. \tag{9.89}$$

Thus the minimal length ℓ plays the role of a time in this rate equation for the total domain density.

The equation of motion for the domain length distribution may now be written in terms of the time-like variable ℓ. When the length of the shortest domain increases from ℓ to $\ell + \Delta\ell$, the length distribution $c(x, \ell)$ changes according to

$$c(x, \ell + \Delta\ell) - c(x, \ell)$$
$$= \left[-2\, \frac{c(x, \ell)}{\rho} + \Theta(x - 3\ell) \int_\ell^{x-2\ell} \frac{c(y, \ell)}{\rho}\, \frac{c(x - \ell - y, \ell)}{\rho}\, dy \right] c(\ell, \ell)\Delta\ell. \tag{9.90}$$

The first term on the right accounts for the loss of a domain of length x due to its merging. The factor of 2 arises because a domain of length x can be either to the left or to the right of the minimal domain. The second term accounts for the gain of a domain of length x due to the merging of three domains of lengths ℓ, y, and $x - \ell - y$. The Heaviside step function $\Theta(x - 3\ell)$ imposes the condition that, when the minimal domain has length ℓ, the shortest domain that can be created by merging must have length $\geq 3\ell$. The last factor $c(\ell, \ell)\Delta\ell$ counts the number of merging events that occur as the minimal size increases from ℓ to $\ell + \Delta\ell$. As also found in the cluster dynamics discussed in Section 8.8, a remarkable feature of extremal dynamics is that if the domains are initially uncorrelated, they remain uncorrelated at all times. Merging of domains does not affect their neighbors, nor are domains affected by their neighbors.

---

[16] The resulting coarsening resembles the Swendsen–Wang and Wolff cluster dynamics that were discussed in Section 8.8.

It is now convenient to work with the normalized length distribution $P(x, \ell) \equiv c(x, \ell)/\rho$. Using Eqs (9.89) and (9.90) (after dividing by $\Delta \ell$), the evolution of $P(x, \ell)$ is governed by

$$\frac{dP(x, \ell)}{d\ell} = \Theta(x - 3\ell)P(\ell, \ell) \int_{\ell}^{x-2\ell} P(y, \ell)P(x - \ell - y, \ell)\, dy, \qquad (9.91)$$

which contains only a gain term. Let's use Eq. (9.91) to determine the asymptotic behavior of the domain length distribution. The perpetually growing minimal domain length ℓ defines a scale that we postulate is characteristic of the system in the long-time limit. Thus we assume that $P(x, \ell)$ approaches the scaling form

$$P(x, \ell) \simeq \frac{1}{\ell}\, \Phi\left(x/\ell\right) \qquad (9.92)$$

as $\ell \rightarrow \infty$. The prefactor in Eq. (9.92) is needed to ensure the normalization $\int_1^{\infty} P(x, \ell)\, dx = 1$. As a consequence, the scaling function $\Phi(z)$ must satisfy $\int_1^{\infty} \Phi(z)\, dz = 1$. Substituting the scaling ansatz for $P(x, \ell)$ into (9.91), we find that the scaling function $\Phi(z)$ obeys

$$z\frac{d\Phi(z)}{dz} + \Phi(z) + \Theta(z - 3)\,\Phi(1)\int_1^{z-2} \Phi(y)\,\Phi(z - 1 - y)\, dy = 0. \qquad (9.93)$$

Since this equation involves a convolution, it is helpful to work in the Laplace domain. For this particular example, it is convenient to define the Laplace transform as $\phi(s) = \int_1^{\infty} \Phi(z)e^{-sz}dz$; i.e. with the lower limit of 1 because the convolution has the same lower limit. We now multiply (9.93) by e^{-sz} and integrate by parts to find that the Laplace transform satisfies

$$s\frac{d\phi}{ds} = -\Phi(1)\left(1 - \phi^2\right)e^{-s}, \qquad (9.94)$$

with the boundary condition $\phi(0) = 1$. To determine the constant $\Phi(1)$ (the normalized density of shortest domains), we substitute in the expansion $\phi(s) = 1 + s\phi'(0) + \cdots$ and take the $s = 0$ limit to obtain $\Phi(1) = \frac{1}{2}$.[17] With $\Phi(1) = \frac{1}{2}$, we can rewrite Eq. (9.94) as

$$\frac{d\phi}{dt} = \frac{1}{2}(1 - \phi^2), \qquad \text{with} \qquad dt = \frac{ds}{s}\, e^{-s},$$

and the solution for the Laplace transform is

$$\phi(s) = \tanh\left[\text{Ei}(s)/2\right], \qquad (9.95)$$

where $\text{Ei}(s)$ is the exponential integral $\text{Ei}(s) \equiv \int_s^{\infty}(1/u)\, e^{-u}\, du$.

There are several basic features of extremal dynamics that are worth emphasizing:

1. The equations of motion are soluble. In contrast, the domain length distribution is unknown for the one-dimensional Ising–Glauber model.

---

[17] Consequently, the asymptotic density of the shortest domains is $P(\ell, \ell) = (2\ell)^{-1}$.

2. Extremal dynamics is deterministic. Randomness arises solely from the initial conditions.
3. Correlations are not generated by the dynamics, in contrast with stochastic evolution processes.
4. The scaling analysis effectively resets the length of the shortest domain to one after each merger. This procedure can be viewed as a fixed point of a real-space renormalization.[18]

We now use the foregoing results to determine the average domain length, $\langle x \rangle = \langle z \rangle \ell$. From the small-argument behavior of the Laplace transform, $\phi(s) = 1 - s\langle z \rangle + \mathcal{O}(s^2)$, and the asymptotic properties of the exponential integral[19] we deduce that $\langle x \rangle = 2\exp(\gamma_E)\,\ell$. Notice that the ratio between the average and minimal domain lengths approaches $2\exp(\gamma_E) \doteq 3.562144$, so that there indeed is a single characteristic length scale of the distribution. From the conservation law $\rho\langle x \rangle = 1$, the total domain density is

$$\rho = \frac{\exp(-2\gamma_E)}{2\ell} \ . \tag{9.96}$$

Finally, the small-length tail of the domain distribution can be evaluated directly from (9.93). In the length range $1 < z < 3$, the integral drops out so that the scaling function satisfies $z\Phi'(z) = -\Phi(z)$, subject to the boundary condition $\Phi(1) = 1/2$. This gives $\Phi(z) = (2z)^{-1}$ for $1 \le z \le 3$. Surprisingly, in the range $\ell < x < 3\ell$, the normalized domain length density $P(x,\ell) \to (2x)^{-1}$ is independent of ℓ.
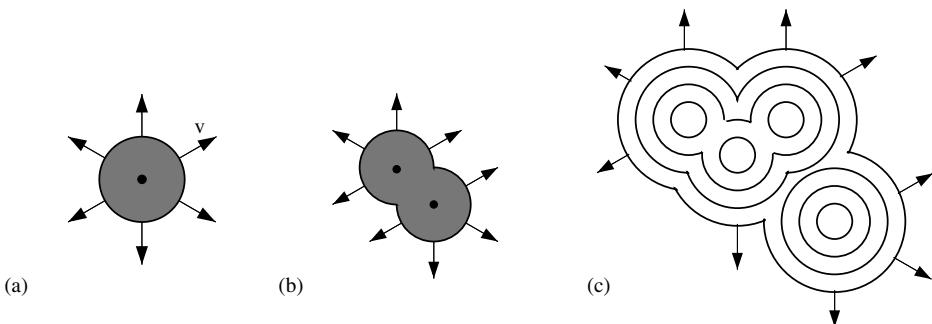
9.8 Nucleation and growth

Phase transitions can be continuous (second order) or discontinuous (first order). Our analysis thus far focused on continuous phase transitions, where we investigated the kinetics when the temperature is quenched from above to below the critical temperature. The same quenching can be used to probe the kinetics of systems exhibiting a discontinuous phase transition. In this section we describe the simple nucleation-and-growth or Kolmogorov–Avrami–Johnson–Mehl (KAJM) model (Fig. 9.11). In nucleation and growth, material in one thermodynamic phase transforms to another phase by the following two steps:

- *Nucleation.* Infinitesimal islands or "seeds" of the stable phase form inside a "sea" of unstable phase.
- *Growth.* From these seeds, stable islands grow at the expense of unstable regions and transform the material into the stable phase. The interface between the stable and unstable phases propagates into the unstable phase at a constant velocity in a direction perpendicular to the interface (Fig. 9.11).

---

[18] A similar *strong disorder renormalization* will be employed in Section 10.2.
[19] For small x, $\mathrm{Ei}\,(x) \simeq -\gamma_E - \ln x$ with $\gamma_E \doteq 0.577215$ the Euler constant.

Fig. 9.11. Illustration of nucleation and growth in two dimensions with initial seeding: (a) one nucleation seed, (b) two nucleation seeds, and (c) four nucleation seeds. The latter is shown at four equally spaced times.

To complete the description of the model, we need to specify how the nucleation seeds of stable phase are actually formed. There are two natural possibilities:

- *Initial seeding.* The system begins with a specified configuration of nucleation seeds.
- *Continuous nucleation.* Seeds nucleate continuously and uniformly in the system at a fixed rate.
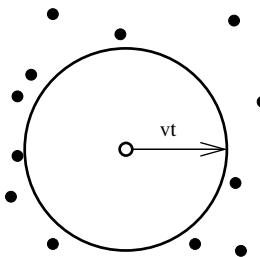
In reality, seeds have a non-zero size that we disregard in the model. The stable phase region therefore consists of the union of a set of growing spherical domains. A basic characteristic of the growth is the rate at which material is transformed into the stable phase. The geometrical nature of the evolution allows us to obtain the fraction of transformed material by simple probabilistic arguments.

Initial seeding

When all seeds nucleate at time $t = 0$ in an unbounded two-dimensional substrate, there are two basic parameters: c, the concentration of nucleation seeds, and v, the propagation velocity from the stable phase into the unstable phase. Since the expansion direction is always perpendicular to the interface, the growing island that emerges from a single circular seed maintains a circular shape of radius vt indefinitely (Fig. 9.11(a)). Multiple isolated seeds eventually coalesce into an expanding region that is the union of circles of radii vt that grow from each seed (Fig. 9.11(b) and (c)). Thus all points that are initially within a distance vt of any nucleation seed are transformed into the stable phase by time t.

For an arbitrary number of seeds, a point in the plane remains in the unstable phase if and only if a circle of radius vt and area $\pi(vt)^2$ is free of nucleation seeds in the initial configuration (Fig. 9.12). Since nucleation is random in space, this circular exclusion zone contains no seeds with probability $\exp[-c\pi(vt)^2]$, which decays exponentially with the expected number of nucleation seeds in this zone, $c\pi(vt)^2$. Hence the area fraction ϕ of the unstable phase is simply

$$\phi(t) = \exp\left[-c\pi(vt)^2\right], \qquad d = 2. \tag{9.97a}$$

Fig. 9.12. In the initial configuration, a point (represented by the small circle) that is more than a distance vt from any seed (dots) will remain in the unstable phase up to time t.

This result is similar in spirit to Eq. (7.47), which gives the area distribution of target zones in irreversible disk adsorption. The area fraction of the stable phase ρ is simply $\rho = 1 - \phi$ and eventually the entire space is covered by the stable phase.

In d dimensions, the shape of the exclusion zone is a ball of radius vt and volume $V_d(vt)^d$, where V_d is the volume of the unit ball in d dimensions.[20] The area fraction of the unstable phase is therefore given by a straightforward generalization of (9.97a):

$$\phi(t) = \exp\left[-cV_d(vt)^d\right]. \tag{9.97b}$$

Again, the system approaches the final state of complete coverage by the stable phase exponentially in time.

Continuous nucleation

For continuous nucleation, seeds are constantly created at rate γ per unit volume starting from an initially empty system. Now the size of an island depends on its age, which is different for each island (Fig. 9.13). Thus we need to keep track of the complete history of the nucleation process. Let's first consider the case of one dimension, where a stable region that emanates from an isolated seed can be viewed as an expanding triangle in a space-time representation (Fig. 9.14(a)). Similarly, stable regions that emanate from multiple nucleation seeds are unions of triangles in this space-time representation. For a point to remain in the unstable phase, nucleation events cannot occur within an inverted triangle of base width $2vt$, height t, and area vt^2 in the x–t plane (Fig. 9.14(b)). The expected number of nucleation events in this exclusion zone is γvt^2. Therefore the fraction of the system that remains in the unstable phase is

$$\phi(t) = \exp\left[-\gamma vt^2\right], \qquad d = 1. \tag{9.98a}$$

Thus continuous nucleation and growth results in a much faster approach to asymptotic behavior than the case of initial seeding and subsequent growth.

---

[20] The unit ball is defined as the set $\{x \mid ||x|| < 1\}$, whereas the unit sphere is the set $\{x \mid ||x|| = 1\}$. The volume of the unit ball in d dimensions is $V_d = \pi^{d/2}/[\Gamma(1 + d/2)]$, with $V_1 = 2$, $V_2 = \pi$, and $V_3 = 4\pi/3$.
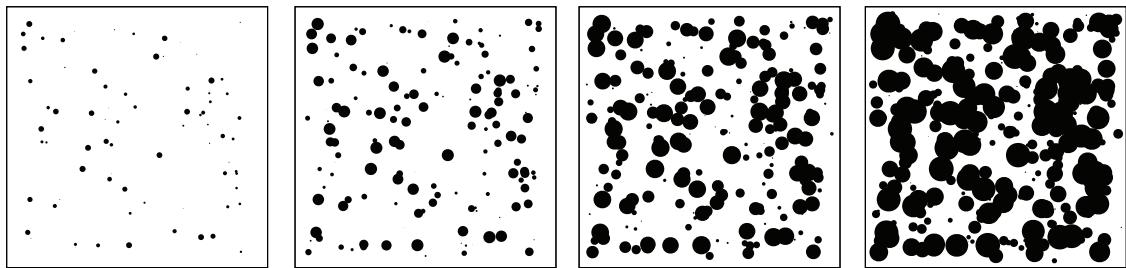
Fig. 9.13.   Snapshots of continuous nucleation and growth for a two-dimensional system at equally spaced times.
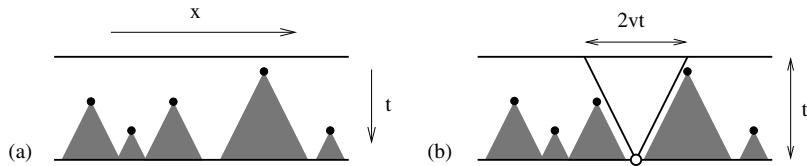
Fig. 9.14.   Nucleation and growth in one dimension with continuous nucleation. (a) Space-time evolution of islands due to seeds that nucleate at different times (dots); the resulting stable region is shown shaded. (b) For the circle to remain in the unstable phase, there must be an inverted triangular exclusion region with base $2vt$ and height t.

The generalization to d dimensions is again straightforward. Now the exclusion zone is a $(d + 1)$-dimensional hypercone with radius vt, height t, and volume

$$\int_0^t V_d \, (v\tau)^d \, d\tau = \frac{1}{d + 1} V_d \, v^d \, t^{d+1}.$$

Consequently, the result of Eq. (9.98a) now becomes

$$\phi(t) = \exp\left[-\frac{1}{d + 1} \gamma \, V_d \, v^d \, t^{d+1}\right], \tag{9.98b}$$

for all d. The above results can also be straightforwardly generalized to account for nucleation rates that depend on time or on space, as well as time-dependent growth velocities.

Spatial correlations

A more complete characterization of the spatial distribution of stable and unstable phases is provided by the two-point correlation function

$$G(x, t) = \langle\varphi(0, t)\varphi(\mathbf{x}, t)\rangle - \langle\varphi(\mathbf{x}, t)\rangle\langle\varphi(0, t)\rangle \qquad \text{with} \qquad x = |\mathbf{x}|, \tag{9.99}$$

which quantifies the probability that two points separated by a distance x are both in the unstable phase at time t. Here $\varphi(\mathbf{x}, t)$ is the phase indicator function, with $\varphi(\mathbf{x}) = 1$ when
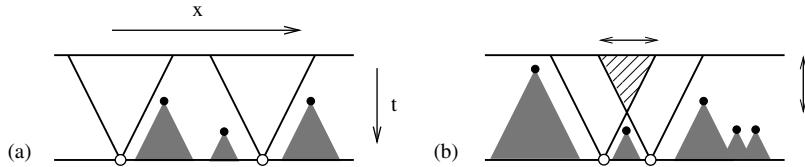
Illustration of the correlation function in one dimension with continuous nucleation for: (a) two uncorrelated points, and (b) two correlated points. The overlap between the exclusion zones (hatched triangle) governs the correlation between the two points.

$\mathbf{x}$ is in the unstable phase and $\varphi(\mathbf{x}) = 0$ otherwise. Thus $\langle \varphi(\mathbf{x}) \rangle = \phi$, the area fraction of the unstable phase. We tacitly assume that the correlation function is isotropic because the system is spatially uniform. By its definition, $\varphi^2 = \varphi$, so that the correlation function G satisfies the boundary condition $G(x = 0) = \phi - \phi^2$. Since $\phi = \langle \varphi(\mathbf{x}) \rangle$ has already been calculated, we need to find

$$g(x, t) = \langle \varphi(0, t) \varphi(\mathbf{x}, t) \rangle, \tag{9.100}$$

from which $G = g - \phi^2$.

Let's first consider continuous nucleation in one dimension. We have already seen that the space-time exclusion zone is an inverted triangle, and that, for a point to remain in the unstable phase, this exclusion zone must be free of nucleation seeds. Similarly, the probability that two points separated by a distance x remain in the unstable phase is governed by the union of two exclusion triangles (Fig. 9.15). If $A_\cup$ is the area of this union, then $g(x) = \exp(-\gamma A_\cup)$, since no nucleation events may occur inside this composite exclusion zone. When $x > 2vt$, we have $A_\cup = 2A$ with $A = vt^2$ the area of one triangle (Fig. 9.15(a)). Using (9.98a), we have $g(x) = \phi^2$ and the two points are therefore uncorrelated, $G(x) = 0$.

However, points that are separated by a distance less than $2vt$ are correlated because there is an overlap between the exclusion zones. The overlap area is a triangle with base $2vt - x$, height $(2vt - x)/2v$, and area (Fig. 9.15(b))

$$A_\cap = \frac{(2vt - x)^2}{4v} . \tag{9.101}$$

Finally, the area of the union of the two exclusion zones is simply $A_\cup = 2A - A_\cap$. Substituting this area into $g(x) = \exp(-\gamma A_\cup)$ gives $g(x) = \phi^{2-F(z)}$, with the normalized overlap function $F(z) \equiv A_\cap/A = (1 - z)^2$. Using (9.99), we obtain the correlation function

$$G(x, t) = \begin{cases} \phi^2 \left[\phi^{-(1-z)^2} - 1 \right], & z \leq 1, \\ 0, & z \geq 1. \end{cases} \quad \text{with} \quad z = \frac{x}{2vt}. \tag{9.102a}$$

A positive correlation exists between points separated by a distance less than $2vt$, and there is no correlation otherwise. The correlation function is a universal function of the coverage ϕ: two different systems at different stages of evolution but at the same overall coverage have identical two-body correlations.
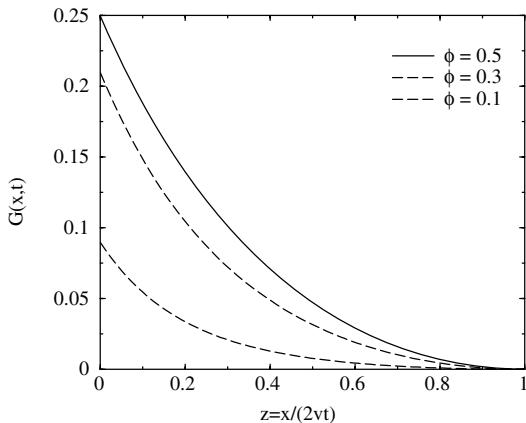
Fig. 9.16.

The scaled correlation function $G(x, t)$ versus scaled distance $z = x/(2vt)$ for the cases $\phi = 0.1$, 0.3, and 0.5.

For arbitrary dimension, the correlation function (9.102a) straightforwardly generalizes to

$$G(x) = \phi^2 \left(\phi^{-F_d(z)} - 1 \right), \tag{9.102b}$$

for $z \leq 1$, with ϕ given by (9.98b). Here $F_d(z)$ is the overlap function in d dimensions that equals the normalized overlap volume between two aligned $(d+1)$-dimensional hypercones of unit height and unit radius that are separated by distance $2z$,

$$F_d(z) = \begin{cases} (1-z)^2, & d = 1, \\ \dfrac{2}{\pi} \left[\cos^{-1} z - 2z\sqrt{1-z^2} + z^3 \ln \dfrac{1 + \sqrt{1-z^2}}{z} \right], & d = 2, \\ (1-z)^3(1+z), & d = 3. \end{cases} \tag{9.103}$$

The overlap function satisfies $F(0) = 1$ and $F(1) = 0$ so that $G(x = 0) = \phi - \phi^2$ and $G(x = 2vt) = 0$. Because the overlap function monotonically decreases with scaled distance z, the correlation function (9.102b) also decreases monotonically with distance, and vanishes for scaled distances $z \geq 1$ (Fig. 9.16). The derivations given above generalize, in a natural way, to treat higher-order correlation functions. For example, to obtain n-point correlations, we must evaluate the volume of the union of n exclusion points. This calculation is the subjects of problems 9.22 and 9.23.

9.9 Notes

General background on coarsening phenomena can be obtained from the two references mentioned in the previous chapter [139, 140], as well as [164, 165]. The Schrödinger potential problems that appear in coarsening are discussed in standard texts, and two useful discussion

are given in Landau and Lifshitz [166] and in Morse and Feshbach [167]. Early work by Cahn and collaborators on coarsening, as well as the Allen–Cahn and Cahn–Hilliard equations is discussed in [140]. The evolution of interfaces by local-curvature flow in which all closed two-dimensional interfaces asymptotically shrink to a circle is presented in [168, 169]. The solution of the evolving finger was given by Mullins [170]. The construction of triply periodic minimal surfaces was discovered by Schwarz [171], while more recent and comprehensive works on this fascinating geometrical subject appear in [172–174]. Derrick's theorem is discussed, for example in [175–177]; Derrick's paper is [178]. The topological properties of defects in systems with a vector order parameter is explored in, for example, [140, 179].

The Lifshitz–Slyozov–Wagner coarsening phenomenon was developed in [180, 181]. The extremal dynamics coarsening of domains in one dimension was investigated and solved in [182–184]. More details about the nucleation-and-growth process can be found in the following two review articles [125, 185].

9.10 Problems

9.1 Examine the equation $m_t = \nabla^4 m$ with the "wrong" positive sign on the right-hand side. Analyze solutions of the form $m(x, t = 0) = \sin kx$. Show that short-wavelength perturbations quickly grow with time.

9.2 Complete the derivations of the autocorrelation function (9.18) and the two-body correlation function (9.19).

9.3 Verify that the kinks for the model with the potential $V = \frac{1}{2} m^2 (1 - m^2)^2$ are given by Eq. (9.26).

9.4 Show that ψ_0 from (9.33) corresponds to the lowest energy state by the following reasoning:

  (a) Verify that $m_0(x)$ is a monotonic function of x.

  (b) Verify that ψ_0 has no nodes (i.e. zeros) due to (a).

  (c) Apply a theorem that, for the one-dimensional Schrödinger equation with an arbitrary potential, the eigenfunction with no nodes has the lowest energy.

9.5 Study the stability of the ground state $m_0 = 1$ for the model with potential $V = \frac{1}{2}(1 - m^2)^2$. Show that the long-time decay of small magnetization perturbations is $\phi \sim e^{-4t}$.

9.6 Study the stability of the kink $m_0 = (1 + e^{-x})^{-1/2}$ for the model with potential $V = \frac{1}{2} m^2 (1 - m^2)^2$.

  (a) Write the corresponding Schrödinger equation.

  (b) Sketch the potential of this Schrödinger equation and show that it monotonically increases from 1 at $x = -\infty$ to 4 at $x = \infty$.

  (c) Using (b) and basic results about the one-dimensional Schrödinger equation, show that there is:

    (i) only one discrete level $E_0 = 0$,

    (ii) a continuous spectrum in the interval $1 < E < 4$, with non-degenerate eigenvalues (each eigenfunction is oscillatory when $x \to -\infty$ and decays as $e^{-x\sqrt{4-E}}$ when $x \to \infty$), and

(iii) finally a continuous spectrum for $E \geq 4$ with doubly degenerate eigenvalues.

9.7 Verify the dependence on parameter a for the terms in the free energy in Eq. (9.51).

9.8 The homotopy group $\pi_i(\mathbb{S}^j)$ classifies mappings of the i-dimensional sphere $\mathbb{S}^i$ into the j-dimensional sphere $\mathbb{S}^j$. This classification does not distinguish mappings that can be continuously deformed one into the other. When $i = j$, every map from $\mathbb{S}^j$ to itself has an integer topological charge measuring how many times the sphere is wrapped around itself. Thus $\pi_j(\mathbb{S}^j) = \mathbb{Z}$ for all $j \geq 1$. Let us analyze the simplest case of mappings of a circle into a circle. The $\mathbb{S}^1 \to \mathbb{S}^1$ mapping can be written as $\theta \to f(\theta)$. The angle θ is defined modulo 2π.

(a) Show that $f(2\pi) - f(0) = 2\pi n$, where n is an integer called the winding number, $w(f) = n$. The winding number plays a role of a topological charge in the case of the circle. This winding number can be rewritten in an integral form

$$n = \frac{1}{2\pi} \int_0^{2\pi} d\theta \, \frac{df}{d\theta} \, .$$

This expression is the analog of Eq. (9.62) which gives the topological charge for the mapping of the usual sphere $\mathbb{S}^2$ into itself.

(b) Show that the mapping

$$f(\theta) = \begin{cases} \theta, & \text{when } 0 \leq \theta \leq \pi, \\ 2\pi - \theta, & \text{when } \pi \leq \theta \leq 2\pi, \end{cases}$$

can be deformed into a trivial mapping $f_0(\theta) = 0$.

(c) Show that the mapping $f_0(\theta) = 0$ cannot be deformed into the mapping $f_1(\theta) = \theta$.

(d) Show that the mapping $f_n(\theta) = n\theta$ cannot be deformed into the mapping $f_m(\theta) = m\theta$ if $n \neq m$.

(e) Consider two mappings $f(\theta)$ and $g(\theta)$ and define their product $f * g$ as the mapping that is initially identical to f but advancing with twice the speed, and then equal to g but advancing at twice the speed. In equations:

$$f * g(\theta) = \begin{cases} f(2\theta), & \text{when } 0 \leq \theta \leq \pi, \\ f(2\pi) + g(2\theta - 2\pi), & \text{when } \pi \leq \theta \leq 2\pi. \end{cases}$$

Show that $f_n * f_m = f_{n+m}$ and that in general the winding numbers add: $w(f * g) = w(f) + w(g)$. This explains that $\pi_1(\mathbb{S}^1) = \mathbb{Z}$ has the structure of the additive group of integers.

9.9 Show that any magnetization that satisfies (9.67), and of course the constraint (9.56), is also a solution of the governing equation (9.58).

9.10 Consider the vortex (9.72).

(a) Verify that it is a solution of Eq. (9.67).

(b) Verify that the topological charge of the vortex (9.72) is $q = 1$.

(c) Compute the free energy of the vortex (9.72) and show that it is equal to 4π.

(d) Construct an explicit expression (similar to (9.72)) for the magnetization distribution of a vortex with charge $q = 2$.

9.11 Consider the *XY* model in two dimensions.

   (a) Show that the constraint (9.56) can be eliminated by employing the representation $\mathbf{m} = (\cos\phi, \sin\phi)$.

   (b) Express the free energy in terms of ϕ and show that the minimization of this free energy yields $\nabla^2\phi = 0$.

   (c) Seek solutions of the equation $\nabla^2\phi = 0$ in the polar coordinates (r, θ). Show that solutions depend only on the angular coordinate, $\phi = \phi(\theta)$, and have the form $\phi = n\theta + \text{const.}$, with n being an integer (the winding number).

   (d) Compute the free energy of the vortex $\phi_n = n\theta$ and show that it is given by

$$F = \frac{1}{2} \int d\mathbf{x} \, (\nabla\phi_n)^2 = \pi n^2 \int_0^\infty \frac{dr}{r}.$$

   This free energy diverges logarithmically (in agreement with our general result that the *XY* model does not admit stationary solutions with finite free energy).[21] The most important vortices are again those with $n = \pm 1$; in the original variables they can be represented as $\mathbf{m} = \pm\mathbf{x}/|\mathbf{x}|$. In hydrodynamic language, the first vortex is a source and the second is a sink. These vortices are singular, not only in the sense that they have diverging energy – there is also a singularity at the origin where the magnetization is ill-defined.

9.12 Consider the "Mexican hat" (or "wine bottle") potential $V = (1 - \mathbf{m}^2)^2$, which is a generalization of the double-well potential $V = (1 - m^2)^2$. Show that there are no non-trivial stationary solutions in any spatial dimension $d \geq 2$, independent of the dimension of the internal space of the order parameter.

9.13 The quasi-static approximation provided the leading asymptotic behavior of the growth law in Eq. (9.81). Estimate the sub-leading correction to the $t^{1/3}$ growth in three dimensions.

9.14 Verify that for the function $g(z)$ defined in (9.82), equation $g(z) = 0$ has a double root (at $z = 3/2$) when $\gamma = \gamma^* = 4/27$.

9.15 Generalize the LSW theory to arbitrary dimension $d > 2$. Verify that the $t^{1/3}$ scaling holds independent of d. Show that the scaling function becomes

$$\phi = C_d z^2 \, (z + 3)^{-1-4d/9} \, (3 - 2z)^{-2-5d/9} \, e^{-d/(3-2z)}$$

   for $z < 3/2$ and $\phi(z) = 0$ for $z \geq 3/2$.

9.16 Consider LSW coarsening for $d = 2$. The absence of an isotropic stationary solution of the Laplace equation that vanishes at infinity leads to a failure of the standard theory. Use the approach developed in Section 2.7 for the reaction rate and show that the result is the slightly slower growth $R_c \sim (t/\ln t)^{1/3}$.

9.17 Complete the derivation of Eq. (9.88).

---

[21] The small size divergence is an artifact of the continuum description; for the microscopic model the lattice spacing provides the lower cutoff. The large size divergence is real and hence for the finite system the free energy of the vortex is proportional to the logarithm of the system size.

9.18 This problem concerns the scaling function $\Phi(z)$ for extremal dynamics that appears in Eq. (9.92).

(a) Compute $\Phi(z)$ in the range $3 \leq z \leq 5$.

(b) Compute $\Phi(z)$ in the range $5 \leq z \leq 7$.

(c) Show that $\Phi(z)$ has discontinuities at the odd integers $z = 1, 3, 5, 7, \ldots$ that become progressively weaker: discontinuity in Φ at $z = 1$, discontinuity in the first derivative Φ' at $z = 3$, discontinuity in the second derivative Φ'' at $z = 5$, etc.

(d) Show that $\Phi(z)$ decays exponentially at the $z \to \infty$ limit, $\Phi(z) \sim Ae^{-az}$. Compute A and a.

9.19 Consider the one-dimensional nucleation-and-growth process.

(a) Let $h(x, t)$ be the density of holes of length x at time t. Show that this hole length distribution evolves according to

$$\left(\frac{\partial}{\partial t} - 2v\frac{\partial}{\partial x}\right) h(x, t) = 2\gamma(t) \int_x^\infty h(y, t)\, dy - \gamma(t)\, x\, h(x, t).$$

(b) Show that the density of holes (or islands) $n(t)$ and the fraction $1 - \rho(t)$ of uncovered space are given by

$$n(t) = \int_0^\infty h(x, t)\, dx, \qquad 1 - \rho(t) = \int_0^\infty x\, h(x, t)\, dx.$$

(c) Show that for instantaneous nucleation, $\gamma(t) = \sigma\delta(t)$, the governing equation turns into the wave equation $h_t - 2vh_x = 0$ whose general solution is $h(x, t) = h_0(x + 2vt)$. Show that the appropriate initial condition is $h_0(x) = \sigma^2 e^{-\sigma x}$ and therefore $h(x, t) = \sigma^2 e^{-\sigma x} e^{-2\sigma vt}$.

(d) Show that for homogeneous nucleation, $\gamma(t) = \gamma$ when $t > 0$, the hole length distribution becomes

$$h(x, t) = (\gamma t)^2\, e^{-\gamma xt}\, e^{-\gamma vt^2}.$$

(e) Let $g(x, t)$ be the density of holes of length x at time t. Show that this distribution obeys

$$\left(\frac{\partial}{\partial t} + 2v\frac{\partial}{\partial x}\right) g(x, t) = \gamma(t)[1 - \rho(t)]\delta(x) + 2v\frac{h(0, t)}{[n(t)]^2}$$
$$\times \left[\int_0^x g(y, t)g(x - y, t)\, dy - 2n(t)g(x, t)\right].$$

9.20 Consider nucleation and growth in one dimension with initial seeding.

(a) Derive the two-point correlation function (9.99).

(b) Write the correlation function in terms of $z = x/(2vt)$ and show that the correlation function depends on coverage and spatial dimension only as in (9.102b).

(c) Obtain the overlap function $F_1(z)$.

9.21 Compute the normalized overlap function $F_d(z)$ for the case of initial seeding of nucleation sites.

9.22 Obtain a formal expression for the n-point correlation function

$$g_n(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n) = \langle \varphi(\mathbf{x}_1)\varphi(\mathbf{x}_2) \cdots \varphi(\mathbf{x}_n) \rangle$$

in d dimensions for nucleation and growth with continuous nucleation at constant rate γ.

9.23 The three-point correlation function is governed by a union of three exclusion zones. Express this union of exclusion zones as a linear combination of single-point exclusion volume, and the volume of overlap between two and three exclusion zones.

Disorder

Order is associated with symmetry, beauty, and simplicity. Some of the most profound advances in physics have exploited these concepts to understand fundamental issues about space and time, as well as the structure of matter at the elemental scale. On the other hand, the notion of disorder resonates with chaos, confusion, and complexity. Such prejudices hindered scientific progress in disordered systems, but by the mid 1970s it was found that many disordered systems exhibit wondrous behaviors that rival the properties of the most symmetrical systems. It was also realized that, in disordered systems, dynamical properties are much more important than in homogeneous systems. In particular, certain disordered systems *cannot* be treated adequately within an equilibrium framework and a dynamical description is necessary.

One prejudice about disordered systems does reflect reality – they can be extraordinarily complicated. There is still considerable opportunity for developing "good" models that capture the essence of how disorder affects a physical phenomenon and yet are either analytically tractable or can be simulated on a reasonable time scale. A prototypical example is the Ising model with non-uniform interactions whose Hamiltonian is

$$\mathcal{H} = -\sum_{\langle ij \rangle} J_{ij} s_i s_j \,. \tag{10.1}$$

The interactions J_{ij} between neighboring spins s_i and s_j are *random* (as well as independent and identically distributed). In a *disordered ferromagnet*, the interactions all have the same sign and the basic question is how the disorder affects the transition to ferromagnetism as the temperature is lowered. When there are ferromagnetic ($J_{ij} > 0$) and antiferromagnetic ($J_{ij} < 0$) interactions with similar overall strengths, the low-temperature phase is more complex. There is no long-range ferromagnetic order at low temperature, but rather *spin-glass* order, in which each spin aligns with its local, but randomly oriented, field (Fig. 10.1). Many aspects of the transition and the spin-glass phase itself still remain mysterious.

The goal of this chapter is to describe a few basic systems where the dynamics is controlled by disorder. These examples give fleeting glimpses into the fascinating world of disordered systems.

10.1 Disordered spin chain

We begin with the Ising chain with disordered nearest-neighbor interactions and with Glauber single spin-flip dynamics. The interactions $J_{i,i+1}$ between neighboring spins are
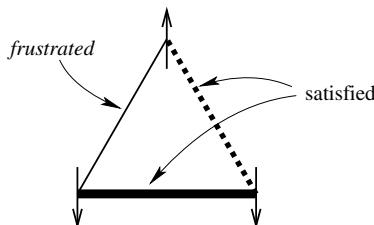
Fig. 10.1. A plaquette in the disordered Ising model (10.1) on the triangular lattice showing ferromagnetic (solid) and antiferromagnetic (dashed) bonds; their thicknesses are proportional to the bond strength. Shown is the minimum energy state where each spin is aligned with the local field due to its two neighbors. The weakest bond is *frustrated* because the orientations of the endpoint spins are opposite to the sense of the interaction.

drawn from a distribution $\rho(J)$. For concreteness, we make three simplifying assumptions about the interaction distribution:

- the $J_{i,i+1}$ are independent identically distributed random variables;
- the distribution $\rho(J)$ is continuous;
- the distribution $\rho(J)$ is symmetric, $\rho(J) = \rho(-J)$.

The most important feature is continuity of $\rho(J)$ so that it is impossible to have two neighboring interactions with the same strength. This restriction ensures that every spin flip either increases or decreases the energy; energy-conserving spin-flip events cannot occur. When the system evolves by Glauber dynamics at zero temperature, energy-raising spin flips are forbidden and only energy-lowering flips are allowed. After such an event, a reversal of the spin cannot occur unless some of its neighbors flip so that the reversal leads to a decrease in the energy of the system.

As in the homogeneous Ising chain, a phase transition occurs at zero temperature and we therefore focus on the zero-temperature evolution. In contrast to the homogeneous ferromagnetic spin chain that has two trivial ground states, the disordered Ising chain has a huge number of metastable states whose number grows exponentially with system size. We shall see that the system gets trapped in one of these metastable states rather than reaching the ground state. While one might anticipate a complex time evolution as a result, some dynamical features are remarkably simple. In particular, an initially aligned system ultimately relaxes to a state where the magnetization equals $\frac{1}{3}$, *independent* of all other system details.

The existence of the huge number of metastable states does not guarantee that the system falls into such a state – the basin of attraction of ground states can in principle dominate and overwhelm the preponderance of metastable states. For example, we saw in Chapter 8 that the homogeneous Ising–Glauber model in two dimensions does not necessarily fall into a metastable stripe state, even though the number of such states is huge (it scales as $[(\sqrt{5}+1)/2]^{\sqrt{N}}$ for a square system of N spins). Nevertheless, it is often the case that a system does fall into a metastable state when the number of such states scales exponentially with system size.
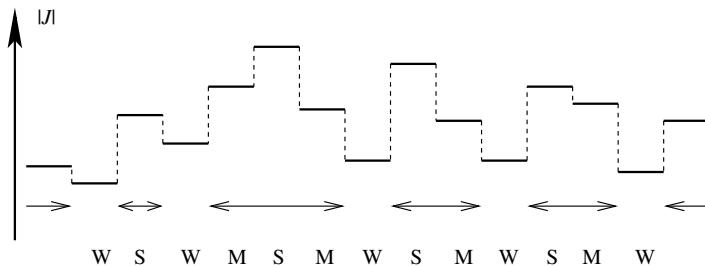
Fig. 10.2. The landscape of bond strengths $|J_i|$ and the corresponding three-letter alphabet – W,M,S – for weak, medium, and strong bonds. The double arrows demarcate the metastable-state clusters, within which spins are aligned, while spins in different clusters are uncorrelated.

Metastable states

In a metastable state, each spin is aligned with its local field. If such a state is reached by zero-temperature Glauber dynamics there is no way to leave this state. The value of the ith spin in a metastable state is determined by

$$s_i = \text{sgn}(J_{i-1}s_{i-1} + J_i s_{i+1}), \tag{10.2}$$

where $J_i \equiv J_{i,i+1}$. Because of the continuity of the interaction distribution, $|J_{i-1}| \neq |J_i|$ with probability 1, and only the stronger of the two interactions determines s_i. For example, if $|J_i| > |J_{i-1}|$, then (10.2) gives $s_i = \text{sgn}(J_i s_{i+1})$, so that s_i is perfectly coupled to s_{i+1}.

A useful way to characterize the metastable states is by the relative strengths of adjacent bonds. We term the ith bond as "strong" if $|J_i| > \max(|J_{i-1}|, |J_{i+1}|)$, and "weak" if $|J_i| < \min(|J_{i-1}|, |J_{i+1}|)$. Otherwise, the bond is "medium." The actual values of the bond strengths are irrelevant for the evolution; all that matters is whether a bond is weak, medium, or strong. Therefore a bond configuration can be represented as the three-letter sequence, W,M,S (Fig. 10.2).

Random multiplicative processes and exponential averages

Consider the product of N random variables in which each factor in the product can be either 2 or 1/2 with probability 1/2. What is the average value of this product, $\langle \Pi \rangle$? Since k factors of 2 and $N - k$ factors of 1/2 occur with probability $2^{-N}\binom{N}{k}$, the average value of the product is

$$\langle \Pi \rangle = 2^{-N} \sum_{k=0}^{N} \binom{N}{k} 2^k (1/2)^{N-k}$$

$$= 2^{-N}(2 + 1/2)^N = (5/4)^N = \langle e^{\ln \Pi} \rangle. \tag{10.3}$$

The exponentially large average value of $\langle \Pi \rangle$ runs counter to one's first intuition that the product should consist of an equal number of factors of 2 and 1/2, and thus equal 1. In fact,

the foregoing gives the *most probable* value of the product, Π_{mp}. It is instructive to write Π_{mp} as

$$\Pi_{mp} = 2^{N/2} (1/2)^{N/2}$$
$$= e^{[\ln 2 + \ln(1/2)]N/2} = e^{\langle \ln \Pi \rangle} = 1 . \qquad (10.4)$$

The logarithm of the product is the sum of N factors, each of which is either $\ln 2$ or $\ln(1/2)$, and each occurs with probability $1/2$. The average value of $\ln \Pi$ is then N times the average of these two factors. We have found the basic result

$$\langle e^{\ln \Pi} \rangle \gg e^{\langle \ln \Pi \rangle} . \qquad (10.5)$$

Why, intuitively, should the mean and most probable value of the product be so different? A typical term in the product does not consist of the same number of 2s and 1/2s, but rather each factor occurs $N/2 \pm \sqrt{N}$ times. Thus a typical term in the sum is of the order of

$$2^{N/2 \pm \sqrt{N}} (1/2)^{N/2 \mp \sqrt{N}} = 2^{\pm 2\sqrt{N}} ,$$

which is already quite different from the most probable value. In fact, the largest term in the sum (10.3) asymptotically grows as $(5/\sqrt{8\pi N}) \times (5/4)^N$, which indicates that rare events make a substantial contribution to $\langle \Pi \rangle$, and cause $\langle \Pi \rangle$ to be different from Π_{mp}.

We define the part of the chain between two weak bonds as a cluster. By definition, weak bonds cannot be adjacent, so each cluster contains at least one bond. There is exactly one strong bond in a cluster and neighboring clusters are separated by single weak bonds. At zero temperature, the state in which every non-weak bond is *satisfied*, that is, $s_i = s_{i+1}$ if the ith bond is ferromagnetic and $s_i = -s_{i+1}$ if the bond is antiferromagnetic, minimizes the energy. For a weak bond, however, the end spins are unrelated and the bond could equally well be satisfied or *frustrated*, where the orientation of adjacent spins is opposite to that of their interaction. Since the state of each weak bond may be specified independently, the total number of metastable states is

$$\mathcal{M} = 2^{\Omega} , \qquad \Omega = \text{number of weak bonds}. \qquad (10.6)$$

To determine the average number of weak bonds, consider three consecutive bonds. Clearly, the middle bond is the weakest with probability $\frac{1}{3}$, so that the average number of weak bonds is $\langle \Omega \rangle = N/3$. What does it tell us about the average number of metastable states? The naive guess that it is given by $2^{\langle \Omega \rangle}$ is wrong. This discrepancy arises because $\mathcal{M}$ grows exponentially with Ω. Even though the statistical properties of Ω may be well behaved, the statistical properties of an exponential function of Ω are not, so that $2^{\langle \Omega \rangle}$ can be very different from $\langle 2^{\Omega} \rangle$ (see the highlight above). Thus we have to carefully distinguish between

the *most probable* number of metastable states (which is given by the above guess)

$$\mathcal{M}_{\mathrm{mp}} = 2^{\langle \Omega \rangle} = 2^{N/3} , \tag{10.7a}$$

and the *average* number of metastable states

$$\langle \mathcal{M} \rangle = \langle 2^{\Omega} \rangle , \tag{10.7b}$$

which we will show grows as $(4/\pi)^N$.

To obtain a complete understanding of the number of metastable states, we need to determine the probability $\Pi_N(\Omega)$ that a chain of N spins contains Ω weak bonds. This can be done by a recursive method that is equivalent to the transfer matrix technique of equilibrium statistical mechanics.

The distribution $\Pi_N(\Omega)$

The idea of the transfer matrix method is to write and solve a recurrence for an N-dependent statistical quantity.[1] In the present case we want to express Π_{N+1} in terms of Π_N. However Π_N alone is inadequate to construct the recurrence. We need to know the strength of the last bond and also whether the previous bond is stronger or weaker than the last bond. To keep track of this information about adjacent bond strengths, we introduce $P_N(\Omega, x)$, the probability to have Ω weak bonds in a system of N bonds, with the conditions $|J_{N-1}| = x$ and $|J_{N-2}| > x$. Likewise, $Q_N(\Omega, x)$ is defined as the probability to have Ω weak bonds, with the conditions $|J_{N-1}| = x$ and $|J_{N-2}| < x$. These probabilities obey the recursion relations
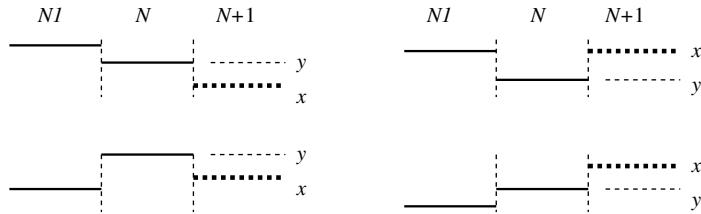
$$\begin{aligned} P_{N+1}(\Omega, x) &= \psi(x) \int_x^\infty [P_N(\Omega, y) + Q_N(\Omega, y)] \, dy, \\ Q_{N+1}(\Omega, x) &= \psi(x) \int_0^x [P_N(\Omega - 1, y) + Q_N(\Omega, y)] \, dy, \end{aligned} \tag{10.8}$$

where $\psi(x) = \rho(x) + \rho(-x)$ is the probability density that a bond strength[2] has absolute value equal to x. The terms in these recursion relations are illustrated in Fig. 10.3. For example, the probability $P_{N+1}(\Omega, x)$ can be built from either $P_N(\Omega, y)$ or $Q_N(\Omega, y)$ in which the strength y of the previous bond exceeds x. Similarly, $Q_{N+1}(\Omega, x)$ can be built from either $P_N(\Omega - 1, y)$ or $Q_N(\Omega, y)$ in which y is less than x. The number of weak bonds increases only by adding a bond of strength $x > y$ to the set of configurations in $P_N(\Omega - 1, y)$ (upper right in Fig. 10.3).

Although the following derivation is somewhat technical, each step is conceptually straightforward and the final result given in Eq. (10.14) is simple. To solve the recursions

---

[1] For example, the Fibonacci numbers are defined by the scalar recurrence $F_{N+1} = F_N + F_{N-1}$. When F_N is a vector quantity, the recurrence involves the transfer matrix.

[2] For the symmetric distribution of the coupling strength, $\rho(x) = \rho(-x)$ and therefore $\psi(x) = 2\rho(x)$. The symmetry has no effect on the "weak bond" concept and therefore this derivation holds for non-symmetric distributions $\rho(x)$. However, the expression $\mathcal{M} = 2^\Omega$ for the number of metastable states does assume symmetry.

Schematic illustration of the four terms in Eq. (10.8). The cartoons on the left correspond to the contributions to $P_{N+1}(\Omega, x)$ (last bond weaker than the previous), while those on the right correspond to the contributions to $Q_{N+1}(\Omega, x)$ (last bond stronger than the previous).

(10.8), we introduce the generating functions

$$\mathcal{P}_N(\omega, x) = \sum_{\Omega \geq 0} e^{\omega \Omega} P_N(\Omega, x), \qquad \mathcal{Q}_N(\omega, x) = \sum_{\Omega \geq 0} e^{\omega \Omega} Q_N(\Omega, x),$$

to recast (10.8) into

$$\mathcal{P}_{N+1}(\omega, x) = \psi(x) \int_x^\infty \left[\mathcal{P}_N(\omega, y) + \mathcal{Q}_N(\omega, y) \right] dy,$$

$$\mathcal{P}_{N+1}(\omega, x) = \psi(x) \int_0^x \left[e^\omega \mathcal{P}_N(\omega, y) + \mathcal{Q}_N(\omega, y) \right] dy. \tag{10.9}$$

Thus $(\mathcal{P}_{N+1}, \mathcal{Q}_{N+1})$ is related to $(\mathcal{P}_N, \mathcal{Q}_N)$ by a linear operator. Denote by e^λ its largest eigenvalue. The quantity $\lambda = \lambda(\omega)$ is known as the *Lyapunov exponent*. In the limit $N \to \infty$, the solutions to Eqs (10.9) grow exponentially with N: $\mathcal{P}_N(\omega, x) \sim p(\omega, x) e^{\lambda N}$ and $\mathcal{Q}_N(\omega, x) \sim q(\omega, x) e^{\lambda N}$. Substituting these asymptotic forms into (10.9) yields

$$e^\lambda p(\omega, x) = \psi(x) \int_x^\infty \left[p(\omega, y) + q(\omega, y) \right] dy,$$

$$e^\lambda q(\omega, x) = \psi(x) \int_0^x \left[e^\omega p(\omega, y) + q(\omega, y) \right] dy.$$

Dividing these equations by $\psi(x)$ and differentiating with respect to x gives

$$e^\lambda \left(\frac{p}{\psi} \right)' = -p - q, \qquad e^\lambda \left(\frac{q}{\psi} \right)' = e^\omega p + q. \tag{10.10}$$

We now introduce the auxiliary variable $dz = \psi(x)\, dx$ and rescale the amplitude function as $\tilde{p}(\omega, z) = p(\omega, x)/\psi(x)$ and $\tilde{q}(\omega, z) = q(\omega, x)/\psi(x)$, so that Eqs (10.10) become

$$e^\lambda \frac{d\tilde{p}}{dz} = -\tilde{p} - \tilde{q}, \qquad e^\lambda \frac{d\tilde{q}}{dz} = e^\omega \tilde{p} + \tilde{q}.$$

Eliminating $\tilde{q}(\omega, z)$ from this pair of equations gives

$$e^{2\lambda} \frac{d^2 \tilde{p}}{dz^2} + (e^\omega - 1) \tilde{p} = 0. \tag{10.11}$$

By the definition of $Q_N(x)$, we have $Q_N(x = 0) = 0$. This gives $q(\omega, 0) = 0$, as well as $\tilde{q}(\omega, 0) = 0$. Similarly we have $p(\omega, \infty) = 0$ leading to $\tilde{p}(\omega, 1) = 0$.[3] A solution of Eq. (10.11) that satisfies $\tilde{p}(\omega, 1) = 0$ is

$$\tilde{p}(\omega, z) = C(\omega) \sin\left[(1 - z) e^{-\lambda} \sqrt{e^\omega - 1}\right],$$

from which

$$\tilde{q}(\omega, z) = C(\omega) \sqrt{e^\omega - 1} \cos\left[(1 - z) e^{-\lambda} \sqrt{e^\omega - 1}\right]$$
$$- C(\omega) \sin\left[(1 - z) e^{-\lambda} \sqrt{e^\omega - 1}\right].$$

The boundary condition $\tilde{q}(\omega, 0) = 0$ gives the Lyapunov exponent

$$\lambda(\omega) = \ln \frac{\sqrt{e^\omega - 1}}{\tan^{-1} \sqrt{e^\omega - 1}}. \tag{10.12}$$

Having computed $P_N(\Omega, x)$ and $Q_N(\Omega, x)$, we can find the distribution $\Pi_N(\Omega)$. Since $P_N(\Omega, x) + Q_N(\Omega, x)$ is the probability to have Ω weak bonds with prescribed last bond $|J_{N-1}| = x$, this sum, when integrated over all x, gives the probability to have Ω weak bonds in a chain of N total bonds:

$$\Pi_N(\Omega) = \int_0^\infty \psi(x) \left[P_N(\Omega, x) + Q_N(\Omega, x)\right] dx.$$

The corresponding generating function $G_N(\omega) = \sum_{\Omega \geq 0} e^{\omega \Omega} \Pi_N(\Omega) = \langle e^{\omega \Omega} \rangle$ is therefore

$$G_N(\omega) \sim e^{\lambda(\omega)N} \int_0^\infty \psi(x) \left[p(\omega, x) + q(\omega, x)\right] dx$$
$$= e^{\lambda(\omega)N} \int_0^1 \left[\tilde{p}(\omega, z) + \tilde{q}(\omega, z)\right] dz$$
$$= e^{\lambda(\omega)N} C(\omega) e^\lambda \sin\left[e^{-\lambda(\omega)} \sqrt{e^\omega - 1}\right]. \tag{10.13}$$

Average number of metastable states

Using $G_N(\omega) = \sum_{\Omega \geq 0} e^{\omega \Omega} \Pi_N(\Omega) = \langle e^{\omega \Omega} \rangle$ and its explicit expression (10.13), we can extract the average number of metastable states. Indeed, since $\langle \mathcal{M} \rangle = \langle 2^\Omega \rangle = \langle e^{\Omega \ln 2} \rangle$, the average number of metastable states is equal to $G_N(\ln 2)$. Therefore from (10.12), we have

$$\langle \mathcal{M} \rangle \sim (4/\pi)^N. \tag{10.14}$$

More generally

$$\lim_{N \to \infty} \frac{1}{N} \ln \langle \mathcal{M}^n \rangle = \lambda(n \ln 2) = \ln \frac{\sqrt{2^n - 1}}{\tan^{-1} \sqrt{2^n - 1}}$$

---

[3] According to the definitions of $\tilde{p}$, $\tilde{q}$, and ψ, the value $x = \infty$ corresponds to $z = \int_0^\infty \psi(y) \, dy = \int_{-\infty}^\infty \rho(y) \, dy = 1$.

for arbitrary n. The small-n behavior allows us to compute the most probable number of metastable states. Indeed,

$$\sum_{\Omega \geq 0} 2^{n\Omega} \, \Pi_N(\Omega) \to (\mathcal{M}_{\mathrm{mp}})^n \qquad \text{as} \quad n \to 0,$$

while on the other hand $(\mathcal{M}_{\mathrm{mp}})^n$ grows as $e^{\lambda(n \ln 2)N}$. Hence

$$\lim_{N \to \infty} \frac{1}{N} \ln \mathcal{M}_{\mathrm{mp}} = \lim_{n \to 0} \frac{\lambda(n \ln 2)}{n} = \frac{1}{3} \ln 2,$$

thereby confirming (10.7a). When N is large, the average number of metastable states greatly exceeds the most probable number:

$$\frac{\langle \mathcal{M} \rangle}{\mathcal{M}_{\mathrm{mp}}} \sim C^N, \quad C = \frac{2^{5/3}}{\pi} \doteq 1.01057.$$

Thus the number of metastable states exhibits huge fluctuations from realization to realization. Occasionally this number vastly exceeds the most probable number, so that the average number of metastable states is larger than the most probable number.

Remanent magnetization

An important characteristic of a spin glass is the *remanent magnetization* that arises by preparing a laboratory spin glass in highly magnetized state at a temperature T, typically by applying a large external magnetic field, and then switching the field off at fixed temperature and waiting until a final state is reached. The final-state remanent magnetization, $m_\infty(T) \equiv m(T, t = \infty)$, probes the structure of the phase space and provides information about the metastable states of the system. These metastable states are extraordinarily complex for a real spin glass, but are more easily characterized in the disordered Ising chain because, as discussed above, the chain breaks up into non-interacting clusters that are delimited by weak bonds.

Let's begin by determining the zero-temperature remanent magnetization for a symmetric interaction distribution $\rho(J)$. Consider any spin that is not attached to a strong bond. Each such spin is perfectly correlated with its neighbor that is at the other end of the stronger of its two attached bonds. For a symmetric distribution of bond strengths, each such spin is therefore equally likely to be up or down, so that it contributes zero to the remanent magnetization. Only spins at the end of strong bonds are relevant. If a strong bond is ferromagnetic, its two end spins point up, while if the bond is antiferromagnetic, these spins are antiparallel. For a symmetric interaction distribution, ferromagnetic and antiferromagnetic couplings are equiprobable, so the contribution to the remanent magnetization equals $(2 + 0)/2$ times the fraction of strong bonds. This leads to

$$m_\infty = \frac{1}{3}.$$

Hence the remanent magnetization is independent of the interaction distribution, as long as it is symmetric and continuous.

Let us now determine the spin evolution by zero-temperature Glauber dynamics. Consider a cluster of ℓ spins $s_1, \ldots, s_\ell$, with a strong bond at $(j, j+1)$. In an infinitesimal time interval dt, a spin s_i to the left of the strong bond ($1 \leq i \leq j$) evolves as

$$s_i(t + dt) = \begin{cases} s_i(t), & \text{probability } 1 - dt, \\ \epsilon_i s_{i+1}(t), & \text{probability } dt, \end{cases} \tag{10.15}$$

where $\epsilon_i \equiv \text{sgn}(J_i)$. The second line in Eq. (10.15) accounts for s_i becoming equal to s_{i+1} in an update if there is a ferromagnetic interaction between the two spins, while $s_i = -s_{i+1}$ if there is an antiferromagnetic interaction. Thus $S_i(t) \equiv \langle s_i(t) \rangle$ satisfies

$$\frac{dS_i}{dt} = -S_i + \epsilon_i S_{i+1}, \qquad 1 \leq i \leq j. \tag{10.16a}$$

Similarly, spins to the right of the strong bond ($j + 1 \leq i \leq \ell$) evolve according to

$$\frac{dS_i}{dt} = -S_i + \epsilon_{i-1} S_{i-1}, \qquad j + 1 \leq i \leq \ell. \tag{10.16b}$$

To determine the evolution of the macroscopic system, we need to average over the distribution of interactions. Since the interactions are independent identically distributed random variables, the ϵ_i are also independent identically distributed variables that take the values ± 1 only. For a symmetric interaction distribution $\rho(J_i)$, then $\epsilon_i = \pm 1$ independently and with equal probability.

Let $\overline{(\ldots)}$ denote the average of a quantity $(\ldots)$ over the distribution of interactions. Thus $\overline{\epsilon} = 0$ by definition. Instead of solving Eqs (10.16a) and then averaging the solutions over the disorder, we first average these equations and then solve. The averaging of linear terms gives $S_i \to \overline{S_i}$, and averaging the quadratic terms is also simple since the factors in products are uncorrelated; for example, $\overline{\epsilon_i S_{i+1}} = \overline{\epsilon_i}\, \overline{S_{i+1}} = 0$ for $i < j$. Thus S_i depends on ϵ_i and S_{i+1}, but in turn S_{i+1} depends on ϵ_{i+1} and S_{i+2}, etc. Therefore S_i depends on $\epsilon_{i'}$ with indices $i' \geq i$. Because the different ϵ are independent, we conclude that ϵ_i and S_{i+1} are uncorrelated in Eq. (10.16aa). Thus averaging (10.16a) and (10.16b), we obtain

$$\frac{d}{dt} \overline{S_i} = -\overline{S_i}, \qquad \text{for all } i \neq j, j+1,$$

with solution $\overline{S_i}(t) = \overline{S_i}(0)\, e^{-t}$.

For the two spins attached to the strong bond $(j, j+1)$, the above procedure does not quite work because s_{j+1} is now determined by its left neighbor s_j. We can, however, directly solve the original equations

$$\frac{dS_j}{dt} = -S_j + \epsilon_j S_{j+1}, \qquad \frac{dS_{j+1}}{dt} = -S_{j+1} + \epsilon_j S_j, \tag{10.17}$$

and then average. By adding and subtracting the two equations in (10.17) we ultimately obtain $S_j = S_{j+1} = \exp\big[(\epsilon_j - 1)t\big]$ for the ferromagnetic initial state: $s_i(0) = 1$ for all i.

Averaging over disorder, we find that the average spin values within a cluster are given by $\overline{S_i} = e^{-t}$ for $i \neq j, j+1$, while for the spins attached to a strong bond $\overline{S_j} = \overline{S_{j+1}} = \frac{1}{2}(1 + e^{-2t})$. In a cluster of ℓ spins, the two spins at the ends of the strong bond evolve according to the latter formula and the remaining $\ell - 2$ spins relax exponentially. Let X_ℓ be the fraction of clusters that contain ℓ spins. The average magnetization is then given by

$$m(t) = \left(1 + e^{-2t}\right) \sum_{\ell \geq 2} X_\ell + e^{-t} \sum_{\ell \geq 2} (\ell - 2) X_l. \tag{10.18}$$

Notice that $\sum_{\ell \geq 2} X_\ell$ is the density of clusters, which equals the weak bond density of $\frac{1}{3}$, since there is a one-to-one correspondence between clusters and weak bonds. Additionally, $\sum_{\ell \geq 2} \ell X_\ell$ equals the fraction of the system that is covered by a cluster of any size, and thus equals 1. With these results, the magnetization reduces to

$$m(t) = \frac{1}{3} \left(1 + e^{-t} + e^{-2t}\right).$$

Thus not only is the remanent magnetization equal to $\frac{1}{3}$, independent of the distribution of interaction strengths, the evolution to this final state is universal.

Let us now compute the final energy per spin $\mathcal{E}$. Each weak bond is satisfied or frustrated equiprobably, so their contribution to the energy vanishes. All non-weak bonds are satisfied and the fraction of such bonds is $\frac{2}{3}$ leading to $\mathcal{E} = -\frac{2}{3} \overline{|J|_{nw}}$. Since $\frac{2}{3}\overline{|J|_{nw}} + \frac{1}{3}\overline{|J|_w} = \overline{|J|}$, where the subscripts nw and w refer to non-weak and weak bonds, respectively, the energy of the metastable state is

$$\mathcal{E}_{\mathcal{M}} = -\overline{|J|} + \frac{1}{3}\overline{|J|_w}, \tag{10.19}$$

which is larger than the ground-state energy $\mathcal{E}_{GS} = -\overline{|J|}$. To evaluate $\mathcal{E}_{\mathcal{M}}$, let $\psi(x)$ be the probability density for a bond coupling strength to have absolute value $x = |J|$. Then the probability density for a weak bond of strength x is given by $\psi(x) \left(\int_x^\infty \psi(y)\, dy\right)^2$. Therefore

$$\mathcal{E}_{\mathcal{M}} - \mathcal{E}_{GS} = \frac{1}{3} \int_0^\infty \left(\int_x^\infty \psi(y)\, dy\right)^2 x\, \psi(x)\, dx, \tag{10.20}$$

so that the energy can be explicitly computed once the interaction distribution is specified (problem 10.2).

10.2 Random walk in a random potential

There are two types of disorder that naturally arise in random walk problems. The first is randomness of the individual steps in the walk that is caused by thermal noise due to collision of the random walker with other particles in the system. These collisions are normally modeled by taking the direction of each step of the walk to be random. As discussed

in Chapter 2, there are well-developed techniques to deal with this randomness and solve for the probability distribution of a random walk. A second source of randomness is that the medium in which the random walker moves is rarely uniform. In many situations, the changes in the medium occur on a much longer scale than the time scale associated with the random walk; hence we can assume that characteristics of the medium vary randomly in space but not in time. Such *quenched* disorder is used to model materials containing impurities, or having intrinsic randomness.

Phenomenology

The Langevin approach provides a simple continuum description of a random walk in a random medium. We start with the Langevin equation $\dot{\mathbf{x}} = \boldsymbol{\xi}$ corresponding to Brownian motion and add a random force that models the medium to obtain

$$\frac{d\mathbf{x}}{dt} = \mathbf{F}(\mathbf{x}) + \boldsymbol{\xi}(t). \tag{10.21}$$

Here $\boldsymbol{\xi}(t)$ is a Gaussian white noise (in time) that models the thermal noise and has the properties

$$\langle \boldsymbol{\xi} \rangle = \mathbf{0}, \qquad \langle \xi_i(t)\, \xi_j(t') \rangle = 2D\delta_{i,j}\, \delta(t - t'). \tag{10.22}$$

The random force term $\mathbf{F}(\mathbf{x})$ is also usually taken to be Gaussian[4] that resembles (10.22), namely

$$\langle \boldsymbol{F} \rangle = \mathbf{0}, \qquad \langle F_i(\mathbf{x})\, F_j(\mathbf{x}') \rangle = 2\Gamma\delta_{i,j}\, \delta(\mathbf{x} - \mathbf{x}'). \tag{10.23}$$
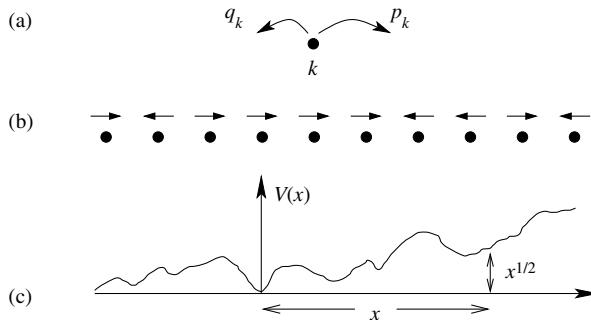
The analysis of (10.21)–(10.23) and related models of the random walk in random media has led to the following remarkably simple characterization:

1. *One dimension*. Quenched disorder significantly slows down the random walk. The typical displacement scales as $(\ln t)^2$ in the case of zero bias.
2. *Two and higher dimensions*. Quenched disorder does not affect the scaling of the displacement with time, but merely affects the diffusion coefficient.[5]

Although it is difficult to prove the validity of the above classification, we can appreciate it heuristically. For a random walk that is discrete in space and time, the number of distinct sites visited by the walker grows as $\sqrt{t}$, $t/\ln t$, t when $d = 1, 2, 3$, respectively. If the disorder is weak, the simplest guess is that it hardly affects the walker. Let us estimate the typical extra displacement $\delta\mathbf{x}$ caused by a random force. We have $\delta\mathbf{x} = \sum_{1 \le \tau \le t} \mathbf{F}[\mathbf{x}(\tau)]$. When $d > 2$, any site is visited at most a few times and therefore there are essentially no correlations in the sequence of forces in the previous sum. Hence the central limit theorem tells us that $\delta x \sim \epsilon\sqrt{t}$, where ϵ is the typical force. This is the same growth as that for the

---

[4] Align (10.23) may sometimes be inadequate; for example, the random force may be incompressible ($\nabla \cdot \mathbf{F} = 0$), or arising from a potential ($\mathbf{F} = -\nabla V$).
[5] For some special random forces, logarithmic corrections arise in two dimensions.

Fig. 10.4. (a) The local hopping bias at site k, with p_k and q_k, the probabilities of hopping to the right and left. (b) The local bias on a string of sites. (c) The random potential that is induced by the local bias.

ordinary random walk, and the factor ϵ tells us that the weak disorder only weakly affects the random walker. This argument makes it plausible that usual diffusive behavior (with a renormalized diffusion constant) should arise above two dimensions. It is also clear that at most logarithmic corrections can arise in two dimensions.

The most interesting phenomena occur in one dimension. First notice that, in one dimension, we can always write the force as the gradient of a potential, $F = -dV/dx$; in particular, we can set $V(0) = 0$ and then define $V(x) = -\int_0^x F(y)\,dy$, as indicated in Fig. 10.4. Since the force itself has zero mean, the random potential also has zero mean and diffusive fluctuations

$$\langle V^2(x) \rangle \sim x\,.$$

The particle thus moves in a random potential. To move a distance L from its initial location, the particle must overcome a potential barrier whose typical height scales as $\sqrt{\Gamma L}$. As we will discuss in the next section, for large L, the random potential may be replaced by a linear potential with the same average gradient. As we learn from problem 2.20, the time required for a diffusing particle to surmount this barrier of height $\sqrt{\Gamma L}$ scales as[6]

$$t \sim t^* e^{\sqrt{\Gamma L}/D}\,, \tag{10.24}$$

where $t^* = \Gamma^2/D^3$. We therefore conclude that the typical length scale of the random walk grows as

$$L \sim \frac{D^2}{\Gamma}\left[\ln(t/t^*)\right]^2\,. \tag{10.25}$$

Thus arbitrarily small quenched randomness drastically affects the motion of a diffusing particle – instead of diffusive growth $L \sim \sqrt{t}$, the particle follows the ultra-slow logarithmic law $L \sim (\ln t)^2$; this behavior is known as *Sinai's law*.

[6] This is a version of the Arrhenius law that states that a particle at a temperature T can overcome a potential barrier V in a time of the order of $e^{V/T}$.

Another remarkable manifestation of the influence of a random potential on diffusive motion arises when two non-interacting particles are subject to the same random potential. If there was no disorder, the positions of the two walkers would be given by

$$\langle x_1^2 \rangle = \langle x_2^2 \rangle = t, \qquad \langle (x_1 - x_2)^2 \rangle = 2t.$$

That is, the distance between the two particles also diffuses, but with twice the diffusion coefficient of the single-particle motion. Disorder completely changes the situation – now the average interparticle distance goes to a finite limit at infinite time! The origin of this unexpected behavior is that each particle tends to move to near the bottom of its local potential well. As time increases, a particle moves to the bottom of progressively deeper valleys. At some point, both particles move to the same valley and then their subsequent positions are strongly correlated because both particles tend to be in the same local valley. This feature is known as the *Golosov phenomenon*.

Decimation procedure

Here we examine the one-dimensional model (10.21)–(10.23), also known as the Sinai model, in more detail. For simplicity, instead of considering the random force (10.23), or equivalently the Brownian-motion random potential, we study a piecewise linear random potential. The slopes of each linear segment have the same absolute value but their signs are random variables, with the same fraction of positive and negative slopes. The asymptotic behavior of a random walker in this random potential is the same as in the Sinai model.
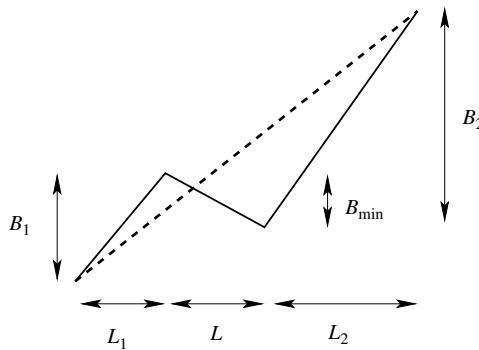
Let ℓ_n be the length of the nth segment along which the potential grows linearly from V_n to V_{n+1}; the endpoints are separated by the barrier $B_n = |V_n - V_{n+1}|$. When the particle moves in such a potential, it will typically remain near the bottom of a valley. If the barrier is small, the particle will soon overcome it and get into a deeper valley. The key point is that, on a long time scale, small barriers play no role. This observation suggests the following recursive procedure that simplifies the random potential but does not affect the asymptotic motion of the particle. At each step, we eliminate the bond $(l, B_{\min})$ with the smallest barrier by merging it with the adjacent bonds (ℓ_1, B_1) and (ℓ_2, B_2). This *decimation* procedure is a version of *real-space renormalization* and is schematically illustrated in Fig. 10.5.

Formally the decimation is defined by

$$(\ell_1, B_1) \oplus (\ell, B_{\min}) \oplus (\ell_2, B_2) \longrightarrow (\ell_1 + l + \ell_2, B_1 - B_{\min} + B_2), \qquad (10.26)$$

and it preserves the zigzag nature of the piecewise linear potential. The feature that makes analysis possible is that correlations do not spontaneously develop by this renormalization: if the barriers are initially uncorrelated, they remain uncorrelated. Equation (10.26) is an example of extremal dynamics and the details of the renormalization transformation resemble those given for coarsening by extremal dynamics in Section 9.7. While the details become somewhat cumbersome, the final result (10.38) for the probability distribution of a particle is compact and has a remarkably simple interpretation.

Let's first disregard the length of the barriers and focus on the distribution of barrier heights $c(B, b)$, defined as the number density of barriers of height B when the minimal

Fig. 10.5. In decimation, the segment with the smallest barrier B_{min} is eliminated by merging with adjacent segments.

barrier height equals b. The master equation for this barrier distribution is

$$\frac{\partial c(B, b)}{\partial b} = c(b, b) \left[-2 \frac{c(B, b)}{\rho(b)} + \int_b^B dB' \, \frac{c(B', b)}{\rho(b)} \frac{c(B - B' + b, b)}{\rho(b)} \right] . \qquad (10.27)$$

Here the minimal barrier height can be viewed as a time-like variable that is computationally more convenient than the time itself. The first term on the right accounts for the loss of a barrier of height B because it is adjacent (on either side) to a minimal barrier. The second term accounts for the gain of a barrier of height B due to the merging of barriers of heights $B', B - B' + b$, and a negative barrier of depth b. Here $\rho(b) = \int_b^\infty dB \, c(B, b)$ is the density of barriers of any height that evolves according to

$$\frac{d\rho(b)}{db} = -2c(b, b) , \qquad (10.28)$$

since three barriers are removed and one is created in each merging event. Using the asymptotically exact Arrhenius estimate $t \sim e^b$, we express the minimal barrier height b as $\ln t$, so that we can re-express final results in terms of the original time variable.

In the $b \to \infty$ limit, we anticipate that the barrier height distribution approaches the scaling form

$$\frac{c(B, b)}{\rho(b)} \simeq b^{-1} \, \Phi(z), \qquad z = \frac{B - b}{b} . \qquad (10.29)$$

Substituting this scaling ansatz into (10.27), the equation for the barrier height distribution simplifies to

$$(1 + z) \frac{d\Phi(z)}{dz} + \Phi(z) + \Phi(0) \int_0^z \Phi(x) \, \Phi(z - x) \, dx = 0 . \qquad (10.30)$$

We can solve Eq. (10.30) by using the Laplace transform to express the convolution as a product. Alternatively, notice that this equation resembles Eq. (5.61) for the scaling function

in constant-kernel aggregation, whose solution is an exponential function. Thus we also seek[7] an exponential solution, $\Phi = a\,e^{-az}$. Substituting this form into Eq. (10.30) we find that a consistent solution is obtained when $a = 1$. Thus

$$\Phi(z) = e^{-z}\,. \tag{10.31}$$

Now using $c(b,b)/\rho(b) = b^{-1}\Phi(0) = b^{-1}$ in conjunction with (10.28), we determine that the total density of barriers is $\rho(b) = b^{-2}$, and that the density of minimal barriers is $c(b,b) = b^{-3}$.

To probe the full barrier distribution, we need $c(\ell, B, b)$, the number density of segments of length ℓ and barrier height B. This joint distribution evolves according to

$$\left(\frac{\partial}{\partial b} + 2\,\frac{c(b,b)}{\rho(b)}\right) c(\ell, B, b) = \int d\ell_1\, d\ell_2\, d\ell_3\, \delta(\ell_1 + \ell_2 + \ell_3 - \ell)$$

$$\times \int dB_1\, dB_2\, \delta(B_1 + B_2 - b - B)\, c(\ell_3, b, b)$$

$$\times \frac{c(\ell_1, B_1, b)}{\rho(b)}\, \frac{c(\ell_2, B_2, b)}{\rho(b)}\,,$$

where the delta functions ensure that the rescaled barrier has the length and height given by the composition rule (10.26). In the $b \to \infty$ limit, the joint distribution acquires the scaling form

$$c(\ell, B, b) = b^{-5}\, \psi(w, z), \qquad \text{with} \qquad w = \frac{\ell}{b^2} \qquad \text{and} \quad z = \frac{B - b}{b}\,. \tag{10.32}$$

This form can be justified by noting that the minimal barrier height b sets the scale for deviation from the minimal height $B - b$, and that the inverse bond density $\rho^{-1} = b^2$ sets the scale for ℓ. By substituting the scaling form (10.32) into the governing equation for $c(\ell, B, b)$ we find that the scaling function satisfies

$$3\psi + (1 + z)\,\frac{\partial\psi}{\partial z} + 2w\,\frac{\partial\psi}{\partial w}$$

$$+ \int dw_1\, dw_2\, dw_3\, \delta(w_1 + w_2 + w_3 - 1) \int dz_1\, dz_2\, \delta(z_1 + z_2 - z)$$

$$\times \psi(w_1, z_1)\psi(w_2, z_2)\psi(w_3, 0) = 0.$$

We now introduce the Laplace transform $\psi(p, z) = \int_0^\infty e^{-pw}\, \psi(w, z)\, dw$ to recast the above equation into

$$\left[1 + (1 + z)\,\frac{\partial}{\partial z} - 2p\,\frac{\partial}{\partial p}\right]\psi(p, z) + \psi(p, 0) \int_0^z \psi(p, x)\, \psi(p, z - x)\, dx = 0\,. \tag{10.33}$$

---

[7] An exponential contains two parameters, $\Phi = A\,e^{-az}$, but we must choose $A = a$ to ensure that $\rho(b) = \int_b^\infty c(B, b)\, dB$.

This equation is a generalization of (10.30), whose solution is exponential (10.31). This fact suggests again seeking an exponential solution

$$\psi(p,z) = \alpha(p)\, e^{-z\,\beta(p)}. \tag{10.34}$$

Substituting (10.34) into (10.33) we find that the exponential ansatz is consistent if

$$2p\frac{d\alpha}{dp} - \alpha = -\alpha\beta, \qquad 2p\frac{d\beta}{dp} - \beta = -\alpha^2. \tag{10.35}$$

By definition, $\psi(p=0,z) = \int_0^\infty dw\,\psi(w,z) = \Phi(z) = e^{-z}$, see (10.31). Therefore $\alpha(0) = \beta(0) = 1$. Equations (10.35) have an elementary form and can be simplified by making the transformations

$$\alpha(p) = qA(q), \qquad \beta(p) = qB(q), \qquad q = \sqrt{p},$$

to convert (10.35) into

$$\frac{dA}{dq} = -AB, \qquad\qquad \frac{dB}{dq} = -A^2.$$

These equations satisfy the conservation law $B^2 - A^2 = $ const., and the solutions are qualitatively different depending on whether the constant is positive, zero, or negative. Only the former case leads to physically acceptable solutions.[8]

The physically relevant solution is $A = (\sinh q)^{-1}$ and $B = \coth q$. In the original variable $q = \sqrt{p}$, we have

$$\alpha(p) = \frac{\sqrt{p}}{\sinh\sqrt{p}}, \qquad \beta(p) = \sqrt{p}\,\frac{\cosh\sqrt{p}}{\sinh\sqrt{p}}. \tag{10.36}$$

Thus $\psi(p,z)$ is given by (10.34) and (10.36), and its inverse Laplace transform gives $\psi(w,z)$. The result is cumbersome, so let's limit ourselves to single-variable distributions. We already know the barrier height distribution $\Phi(z)$, and we also need the scaling form of the barrier length distribution $\Psi(w) = \int \psi(w,z)\,dz$ (similarly, the Laplace transforms are related by $\Psi(p) = \int \psi(p,z)\,dz$). From (10.34) we find

$$\Psi(p) = \frac{\alpha(p)}{\beta(p)} = \frac{1}{\cosh\sqrt{p}}.$$

Strikingly, this result coincides with the Laplace transform for the first-passage probability for a diffusing particle in a finite, absorbing interval. The inverse Laplace transform of the function $\left(\cosh\sqrt{p}\right)^{-1}$ may be determined by standard residue calculus and gives, for the

---

[8] For example, if the constant were zero we would obtain $A = B = A(0)/[1 + qA(0)]$ and therefore $\alpha(p) = \beta(p) = A(0)\sqrt{p}/[1 + A(0)\sqrt{p}]$ which disagrees with the boundary condition $\alpha(0) = \beta(0) = 1$.

scaled length distribution,

$$\Psi(w) = \pi \sum_{n=-\infty}^{\infty} (-1)^n \left(n + \tfrac{1}{2}\right) \exp\left[-\pi^2 w \left(n + \tfrac{1}{2}\right)^2\right] \tag{10.37}$$

$$= \frac{1}{\pi^{1/2} w^{3/2}} \sum_{n=-\infty}^{\infty} (-1)^n \left(n + \tfrac{1}{2}\right) \exp\left[-\frac{1}{w} \left(n + \tfrac{1}{2}\right)^2\right].$$

We give the answer in two equivalent forms that are convenient for extracting the large- and small-w behaviors, respectively.

The foregoing discussion was devoted to renormalizing the random potential. To determine the position of a diffusing particle we use the fundamental observation (outlined at the beginning of this section) that the particle is at the bottom of the renormalized potential well that contains the starting point. Let $\text{Prob}(x, t)$ be the probability density that this particle is at position x at time t under the assumption that it was at $x = 0$ at $t = 0$. We will need this probability when averaged over environments; we denote the average by $\overline{\text{Prob}(x, t)}$. If $P(\ell, b)$ is the probability that a renormalized segment has length ℓ, then the particle will be in such a segment with probability $\ell P(\ell, b) / \int \ell P(\ell, b) \, d\ell$. The bottom of the valley is uniformly distributed in $[0, \ell]$, so all segments of length $\ell \geq |x|$ contribute to $\overline{\text{Prob}(x, t)}$:

$$\overline{\text{Prob}(x, t)} = \left[2 \int_0^{\infty} \ell P(\ell, b) \, d\ell\right]^{-1} \int_{|x|}^{\infty} P(\ell, b) \, d\ell.$$

We know that $P(\ell, b) \simeq b^{-2} \Psi(\ell/b^2)$, with $\Psi(w)$ given by (10.37). Using this fact and recalling that $b = \ln t$, we arrive at

$$\overline{\text{Prob}(x, t)} = (\ln t)^{-2} F(X), \qquad X = \frac{x}{(\ln t)^2},$$

with
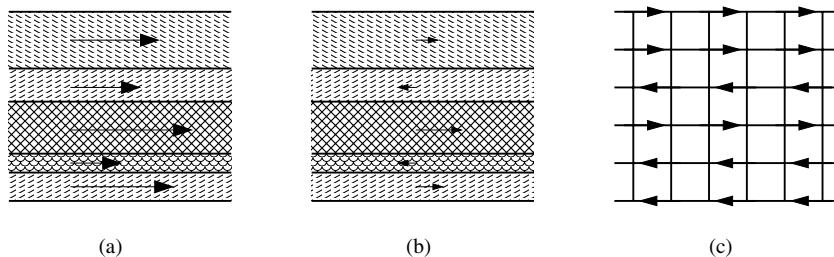
$$F(X) = \frac{2}{\pi} \sum_{n \geq 0} (-1)^n \left(n + \tfrac{1}{2}\right)^{-1} \exp\left[-\pi^2 |X| \left(n + \tfrac{1}{2}\right)^2\right]. \tag{10.38}$$

This is the *Golosov–Kesten* distribution that is the counterpart of the Gaussian distribution for the ordinary random walk. Although the Golosov–Kesten distribution (10.38) is an infinite series, the moments of the displacement (averaged over the noise and all environments) are simple. For example,

$$\overline{\langle x^2 \rangle} \simeq \frac{61}{180} (\ln t)^4, \quad \overline{\langle x^4 \rangle} \simeq \frac{50521}{75600} (\ln t)^8, \quad \overline{\langle x^6 \rangle} \simeq \frac{199360981}{60540480} (\ln t)^{12}. \tag{10.39}$$

10.3 Random walk in random velocity fields

As discussed in the previous section, spatial disorder in the hopping rates of a random walk can lead to the mean-square displacement growing slower than linearly with time.

(a)   (b)   (c)

Fig. 10.6. (a) Flow field in a two-dimensional layered medium. (b) The same flow in a center-of-mass reference frame, from which we abstract the MdM model in (c).

Conversely, there are situations where *superdiffusive* motion can occur, in which the mean-square displacement grows faster than linearly with time. One such example is the Matheron–de Marsily (MdM) model that was formulated to describe groundwater transport in sedimentary rocks.

In the MdM model, a sedimentary medium is modeled as an array of parallel layers, each with slightly different material properties (Fig. 10.6). In each layer, the fluid moves at a constant velocity that is characteristic of that layer. In a reference frame that moves at the average velocity of all the layers, the relative velocity in each layer therefore is transformed into a random, zero-mean function of the transverse position. Now consider the motion of a particle that diffuses and is also passively convected by the flow field. When the particle remains within a given layer, it is convected at the fluid velocity in that layer. Because of superimposed diffusion, however, the particle will make transitions between neighboring layers. Consequently, the longitudinal velocity of the particle will change stochastically. This combination of transverse diffusion and random convection in the longitudinal direction leads to superdiffusion.

A simple approach to understand the particle motion is to consider the Langevin equations of motion. In two dimensions, these equations for a Brownian particle that experiences the additional random horizontal convection are

$$\frac{dx}{dt} = u[y(t)] + \eta_x(t), \qquad \frac{dy}{dt} = \eta_y(t). \tag{10.40}$$

Here $u[y(t)]$ is the velocity field in the x direction that depends only on transverse position. To simplify this system further, we replace the continuum picture in Fig. 10.6(b) by the discrete picture in Fig. 10.6(c) in which the velocity in each layer is randomly $\pm u_0$. By construction, the average longitudinal velocity over all layers equals zero, and there are no correlations between the velocities in different layers. As we shall see, the effect of convection dominates over diffusion so that we may neglect the subdominant effect of the noise η_x in the x direction.

We can determine the time dependence of the mean-square displacement in the longitudinal direction by estimating the magnitude of an effective time-dependent bias that a Brownian particle has experienced up to time t. If every layer in the discrete system of Fig. 10.6(c) was visited with equal frequency, then the average bias would be zero.

However, a diffusing particle typically roams over a perpendicular distance $\sqrt{Dt}$ and thus visits $N \sim \sqrt{Dt}/a$ layers up to time t; here a is the lattice spacing. The average bias velocity within this finite number of layers is

$$\bar{u}(t) \sim \frac{1}{\sqrt{Dt/a^2}} \sum_{y=1}^{\sqrt{Dt/a^2}} u(y) \sim u_0 \left(\frac{a^2}{Dt} \right)^{1/4} . \tag{10.41}$$

Thus the typical longitudinal displacement grows superdiffusively

$$\bar{x} \sim \bar{u}\, t \sim a^{1/2}\, D^{-1/4}\, u_0\, t^{3/4} . \tag{10.42}$$

A more complete characterization of the particle motion is the probability distribution of displacements. When averaged over all particle trajectories and over all realizations of the velocity disorder, we expect that this displacement distribution at long times should have the scaling form

$$P(x,t) \sim t^{-3/4} f(xt^{-3/4}). \tag{10.43}$$

In terms of the scaling variable $\xi \equiv xt^{-3/4}$, the natural expectation is that the scaling function should decay faster than any power law for $\xi \gg 1$; this behavior has been found for essentially all random-walk-like processes. Thus we write

$$f(\xi) \sim \exp(-\text{const.} \times \xi^\delta) . \tag{10.44}$$

We now invoke the general argument that was first presented in the highlight on page 120 to determine the shape exponent δ.

In the case of the MdM model, there is a subtlety in applying this extremal argument because there are two sources of disorder – the disorder in the trajectory and the disorder in the medium. The interplay between these two attributes leads to walks with the largest possible displacement. Clearly, the walk travels furthest longitudinally if it is contained within a subset of layers that all have the same velocity. What is the probability for such an event? In the discretized MdM model in which the velocity for a given value of y is equiprobably $\pm u_0$, the probability that a region of width w consists of layers with the same velocity is 2^{-w}. In the long-time limit, the probability that the transverse position of the walk remains within this region[9] decays as $\exp(-Dt/w^2)$. The confining probability, averaged over all widths, is then

$$\int 2^{-w}\, e^{-Dt/w^2}\, dw \equiv \int e^{F(w)}\, dw , \tag{10.45}$$

with $F(w) = Dt/w^2 - w \ln 2$ which has a single maximum within the integration range. Integrals of the type given in Eq. (10.45) are known as *Laplace integrals* and they can be estimated easily by the *Laplace method*.

[9] This probability that a random walk remains confined to a fixed-size interval also underlies the survival probability of a one-dimensional random walk in the presence of randomly distributed traps, a system that will be discussed in Section 13.2.

Laplace integrals

A Laplace integral has the generic form

$$I(N) = \int_0^\infty e^{NF(x)} \, dx \, ,$$

where $F(x)$ has a single maximum at $0 < x_0 < \infty$. We are interested in the value of this integral in the limit $N \to \infty$. The Laplace method for evaluating this integral involves expanding $F(x)$ about its maximum: $F(x) = F(x_0) - \frac{1}{2}(x - x_0)^2 \, |F''(x_0)| + \cdots$. Using this second-order expansion, we estimate the integral as

$$I(N) = \int_0^\infty \exp\left[N\left(F(x_0) - \frac{1}{2}(x - x_0)^2 \, |F''(x_0)| + \cdots \right)\right] dx$$

$$\simeq \exp[N(F(x_0)] \int_{-\infty}^\infty \exp\left[-\frac{1}{2}(x - x_0)^2 \, |F''(x_0)| \right] dx$$

$$\simeq \sqrt{\frac{2\pi}{N|F''(x_0)|}} \; \exp\left[N|F(x_0)| \right].$$

More precisely, the exponentially large term is the *controlling factor* and the above expression represents the *leading behavior* of the integral; the full expression has this leading behavior multiplied by an infinite asymptotic series in inverse powers of N.

Applying the Laplace method from the above highlight, we find that the optimal value of w scales as $(Dt)^{1/3}$. Consequently, the controlling factor of (10.45) is given by

$$\exp\left[-\text{const.} \times (Dt)^{1/3} \right] .$$

Now using (10.44) for the probability distribution in the MdM model and applying the same comparison for the probability of a stretched walk (as given in the highlight on page 120), we have

$$P(x \sim t, t) \sim t^{-3/4} \exp\left[-\text{const.} \times t^{\delta/4} \right] .$$

Comparing these two expressions we infer that $\delta = 4/3$. A surprising feature of the MdM model is that the shape and distance exponents do not obey the general relation $\delta = (1-v)^{-1}$ (see Eq. (4.41)) that typifies many types of random walk processes.

What happens in the MdM model for the physical case of three dimensions? There are two situations that naturally spring to mind: parallel sedimentary slabs or sedimentary filaments (Fig. 10.7). The former case is perhaps more likely to occur geologically, but it has the same behavior as the two-dimensional system. The latter case is what may be viewed as the true three-dimensional MdM model. Now the equations of motion for a Brownian particle are

$$\frac{dx}{dt} = u[y_1(t), y_2(t)], \qquad \frac{dy_1}{dt} = \eta_1(t), \qquad \frac{dy_2}{dt} = \eta_2(t),$$
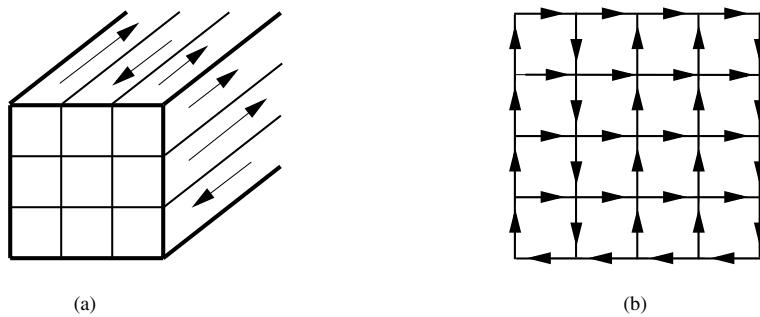
Fig. 10.7. (a) The flow field of the MdM model in three dimensions. (b) The isotropic version of the MdM model – the random Manhattan grid – in two dimensions.

where we view x as the longitudinal coordinate and y_1 and y_2 as the transverse coordinates. Once again, it is helpful to think of the lattice model in which there are random velocities in the x direction that equiprobably take the values $\pm u_0$ for fixed values of the transverse coordinates (y_1, y_2). To determine the longitudinal displacement, it is again helpful to decompose the motion into longitudinal and transverse components. In the transverse direction, the particle trajectory is just a pure random walk in two dimensions. From classic results about random walks in two dimensions, we know that the particle typically visits of the order of $t/\ln t$ distinct sites in the transverse subspace (now ignoring factors that involve D and the lattice spacing). Thus the mean bias that the walk experiences up to time t is given by the extension of Eq. (10.41) to three dimensions. Since the number of distinct filaments visited within time t scales as $t/\ln t$, we have

$$\bar{u} = \frac{1}{t/\ln t} \sum^{t/\ln t} u(y) \sim u_0 \sqrt{\frac{\ln t}{t}}. \qquad (10.46)$$

Then we estimate the typical displacement to be

$$\bar{x} \sim \bar{u}\, t \sim (t \ln t)^{1/2}.$$

The MdM model can be generalized in a natural way to an isotropic random velocity field. An especially interesting realization is the "random Manhattan" lattice, in which the velocity along the entire length of any north–south avenue or east–west street is fixed, but the fractions of one-way north and one-way south avenues, and similarly for east and west streets, are equal. We may obtain the mean-square displacement for a random walk on the random Manhattan lattice by generalizing of the arguments that led to Eqs (10.42). We first arbitrarily partition the isotropic motion of the walk into distinct transverse and longitudinal components, even though the motion is isotropic. Now let's determine the residual longitudinal bias in a typical region that is swept out by the transverse Brownian motion. For the (unknown) displacement in the transverse direction, we write $y \sim t^{\nu}$. Then from Eq. (10.41), the mean longitudinal velocity in the x direction at time t, averaged over the t^{ν} layers that a typical random walk visits during its excursions in the y direction, vanishes as $t^{-\nu/2}$. From the analog of Eq. (10.42), we conclude that $\bar{x} \sim t^{1-\nu/2}$. By isotropy, however,

one must have $\nu = 1 - \nu/2$, or $\nu = 2/3$. Thus a random walk in the random Manhattan lattice is again superdiffusive in which the displacement scales as $t^{2/3}$ (compared to $\bar{x} \sim t^{3/4}$ for the MdM model). The generalization of this argument to a random Manhattan lattice in arbitrary spatial dimension d yields $\nu = 2/(d+1)$. Since the motion can be no slower than pure diffusion, we conclude that, for $d > 3$, $\bar{x} \sim t^{1/2}$. Thus the critical dimension for a random walk in a random Manhattan lattice is $d_c = 3$. For $d > d_c$, $\nu = 1/2$, while for $d < d_c$, $\nu = 2/(d+1)$, and it is natural to expect that logarithmic corrections will arise for $d = d_c$, just as in the MdM model in $d = 3$.

10.4 Notes

Our presentation has focused on a few specific systems where disorder plays a central role in the kinetic behavior and where analytical solutions are possible. The relaxational properties and the remanent magnetization of the disordered kinetic Ising spin chain were given in [186,187]. An elementary pedagogical account of random multiplicative processes, that underlies the counting of the number of weak bonds and expands on the highlight on page 325, is given in [188]. More information about weak bonds and random landscapes (see also problems 10.3–10.5) can be found in [186, 189–191].

It is currently unknown how to compute the remanent magnetization for two- and three-dimensional zero-temperature spin glasses (the analog of the result $m_\infty = \frac{1}{3}$ in one dimension). For more on this intriguing puzzle, see [192]. The remanent magnetization or remanent energy is also unknown for the Sherrington–Kirkpatrick model, which is the simplest non-trivial spin-glass model at zero temperature. The non-equilibrium behavior of the Sherrington–Kirkpatrick model is still poorly understood, and we haven't touched on this beautiful model in this book. The reader is encouraged to read more on spin-glass models from the numerous references contained in [192].

Early work on the Sinai model and other models of random walks in a random potential is reviewed in [193]. The real-space renormalization method for treating these models was developed in [194]; an elementary account of this method appears in [17], Chapter 4. General presentations of random walks in disordered environments and reviews of the Matheron and de Marsily model are given in [14, 15, 195, 196]. Additional information about random walks in random environments can be found in [197, 198]. Basic results about the number of distinct sites visited by a random walk can be found in [1,3]. Finally, an excellent exposition about Laplace integrals is given by Bender and Orszag [138].

10.5 Problems

10.1 This problem concerns the moments of total number of weak bonds.
   (a) Take the explicit formula (10.13) for the generating function $G_N(\omega) = \sum_{\Omega \geq 0} e^{\omega \Omega} \Pi_N(\Omega) = \langle e^{\omega \Omega} \rangle$ and expand it in powers of ω.

(b) Using the above results extract the variance:

$$\langle \Omega^2 \rangle - \langle \Omega \rangle^2 = \frac{2}{45} N \quad \text{as} \quad N \to \infty.$$

(c) Derive the variance directly without using (10.13). Namely, the argument that led to $\langle \Omega \rangle = N/3$ can be generalized to find the variance (and other moments $\langle \Omega \rangle^p$).

10.2 Using (10.19) and (10.20) show that for the uniform distribution of coupling strengths, $\psi(x) = 1$ for $0 \le x \le 1$ and zero otherwise, the ground-state energy is $\mathcal{E}_{GS} = -\frac{1}{2}$, while the metastable-state energy is $\mathcal{E}_M = -\frac{17}{36}$. Similarly, show that for the exponential distribution, $\psi(x) = e^{-x}$, the energies are given by $\mathcal{E}_{GS} = -1$ and $\mathcal{E}_M = -\frac{26}{27}$.

10.3 Consider the cluster size distribution that appears in (10.18). Determine this distribution using the transfer matrix approach. You should get

$$X_\ell = 2^\ell \frac{(\ell - 1)(\ell + 2)}{(\ell + 3)!}.$$

Notice that this distribution is also independent of the distribution of interaction strengths. Verify the validity of the sum rules: $\sum_{\ell \ge 2} X_\ell = 1/3$ and $\sum_{\ell \ge 2} \ell X_\ell = 1$.

10.4 Think of Fig. 10.2 as a one-dimensional landscape: each site has a certain height, the heights are independent identically distributed random variables, and the height distribution is irrelevant as long as it does not contain delta functions. (For computations, it is convenient to draw heights from the uniform distribution on the unit interval $(0, 1)$.) A weak bond corresponds to a valley in the landscape language, and a strong bond corresponds to a peak.

(a) Show that the probability that a sequence of ℓ consecutive heights is ascending is $1/\ell!$.

(v) Show that the probability P_{ij} that $i + j + 1$ consecutive heights contain a valley which is separated by distance i from the peak on the left and distance j from the peak on the right is

$$P_{ij} = \frac{1}{(i + 1)!(j + 1)!} \left[\frac{ij}{\ell + 1} + \frac{1}{\ell + 3} \right].$$

(c) Let X_ℓ be the length distribution between neighboring peaks, or equivalently neighboring valleys. Show that

$$X_\ell = \sum_{i,j \ge 1, i+j=\ell} P_{ij}.$$

Substitute the expression for P_{ij} determined above and compute the sum to give the expression for X_ℓ from the previous problem.

10.5 Consider a two-dimensional landscape in which each site of the square lattice has a certain height, and the distribution of these heights has the same assumptions as the previous problem.
(a) Show that the density of valleys is $1/5$. (The density of peaks is the same.)
(b) Show that the density M_3 of points with three directions of descent is $M_3 = 1/5$. (The density of points with one direction of descent is the same: $M_1 = 1/5$.)
(c) Show that the density M_2 of points with two directions of descent is also $M_2 = 1/5$.
(d) Show that the density of saddles is $1/15$.
(e) Show that the variance of the total number of valleys is

$$\langle \mathcal{V}^2 \rangle - \langle \mathcal{V} \rangle^2 = \frac{13}{225} N \qquad \text{as} \qquad N \to \infty.$$

The average number of valleys is proportional to the density, i.e. $\langle \mathcal{V} \rangle = N/5$.

10.6 Consider the one-dimensional random walk in a random environment that is governed by Eqs (10.21)–(10.23). Repeat the argument that led to (10.25) in a dimensionally correct form and show that

$$L \sim L_* \left(\ln \frac{t}{t_*} \right)^2, \qquad L_* = \frac{D^2}{\Gamma}, \qquad t_* = \frac{\Gamma^2}{D^3}.$$

Explain the appearance of $t_* = \Gamma^2/D^3$ and $L_* = D^2/\Gamma$ using dimensional analysis.

10.7 Consider the one-dimensional random walk in a random environment with an overall bias, namely, instead of (10.23), we have

$$\langle F \rangle = F_0, \qquad \langle F(x) F(x') \rangle - F_0^2 = 2\Gamma \, \delta(x - x').$$

Show that there is one dimensionless combination of the three parameters D, F_0, and Γ, namely

$$\mu = \frac{DF_0}{\Gamma}$$

or any function thereof. It has been found that the random walker moves ballistically when $\mu > 1$,

$$\lim_{t \to \infty} \frac{\langle x \rangle}{t} = V, \qquad V = F_0 \left(1 - \mu^{-1} \right).$$

(Note that the speed V is smaller than the "bare" speed $V_0 = F_0$ which the walker would acquire in the absence of disorder.) When $0 < \mu < 1$, the movement is sub-ballistic: $\langle x \rangle \sim t^\mu$.

10.8 Deduce the amplitudes for the moments given in Eq. (10.39).

11 Hysteresis

Dictionaries (such as *Webster's*) define hysteresis as "a retardation of the effect when the forces acting upon a body are changed; *especially*, a lagging in the values of resulting magnetization in a magnetic material (as iron) due to a changing magnetizing force." Hysteresis originated in studies of magnetism, but it occurs in diverse systems (mechanical, chemical, biological, optical, electronic, etc.). In this chapter, we adopt the perspective of magnetism and seek the response of a spin system to an oscillating magnetic field. We examine how the area and the shape of the hysteresis loop depend on the amplitude and the frequency of the oscillating magnetic field. We utilize ideas and methods described in earlier chapters on kinetic spin models and disordered systems, as well as the tools of dynamical systems and asymptotic techniques.

11.1 Homogeneous ferromagnets

A qualitative understanding of hysteresis already emerges from the behavior of the homogeneous Ising ferromagnet in an oscillating external magnetic field $h(t) = H \sin \Omega t$. We want to determine how the magnetization $m(t)$ depends on the temperature T, as well as on the amplitude H and frequency Ω of the driving field. We will see that the magnetization $m(t)$ lags the field, and after some transient $m(t)$ follows a dynamic $m(t)$–$h(t)$ loop of area

$$A(T, H, \Omega) = \oint m \, dh. \tag{11.1}$$

If this area remains non-vanishing even in the quasi-static limit, $\Omega \to 0$, the system exhibits *static hysteresis*. This traditional form of hysteresis occurs below the ferromagnetic phase transition, $T < T_c$, while above the phase transition, there is no static hysteresis (Fig. 11.1). When the magnetic field is varying non-adiabatically ($\Omega > 0$), the character of the hysteresis loop changes qualitatively.

The area $A(T, H, \Omega)$, or the excess area $A(T, H, \Omega) - A(T, H, 0)$ in the case where the system exhibits static hysteresis, is called the *dynamic hysteresis area*. The dynamic hysteresis area is important in many technological applications. For example, hysteresis measures the additional power dissipated by repetitive switching in magnetic and optical switches. The dynamical hysteresis area has interesting dependences on H and on Ω. When $H \ll 1$, the limiting behaviors are

$$A(T, H, \Omega) \sim \begin{cases} H^a \, \Omega^b, & \text{for } \Omega \ll 1, \\ H^\alpha \, \Omega^{-\beta}, & \text{for } \Omega \gg 1. \end{cases} \tag{11.2}$$
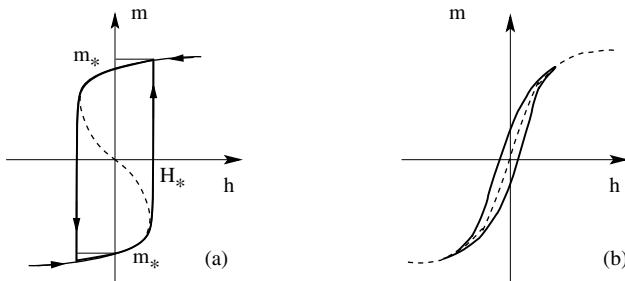
(a) Static hysteresis loop (solid) below the phase transition. The dashed curve represents the equilibrium solution. The system can be trapped in a metastable state with $m < 0$ while $h > 0$, and vice versa. When the field exceeds the coercive field H_*, the system returns to equilibrium, where m and h have the same sign. **(b)** Dynamic hysteresis loop (solid) above the phase transition. In the quasi-static limit the magnetization follows the dashed equilibrium solution $m_{eq}(h)$, and the loop area vanishes.

Let's try to understand these phenomenological laws within the simplest setting, namely, the mean-field ferromagnet in an oscillating magnetic field. The interaction between any pair of spins s_i and s_j is $J_{ij} = 1/N$, and the spins evolve by Glauber dynamics. The Hamiltonian and the spin-flip rate are (see Section 8.4)

$$\mathcal{H} = -\frac{1}{N}\sum_{i<j} s_i s_j - h(t)\sum_{i=1}^{N} s_i, \qquad w_i = \frac{1}{2}\left[1 - s_i \tanh\left(\frac{m + h(t)}{T}\right)\right].$$

From the general relation $\dot{S}_i = -2\langle s_i w_i \rangle$ (see Eq. (8.49a)), the local magnetization $S_i = \langle s_i \rangle$ evolves as

$$\frac{dS_i}{dt} = -S_i + \tanh\left(\frac{m + h(t)}{T}\right).$$

Summing these equations and using $h(t) = H\sin\tau$ to express m in terms of the natural time variable $\tau \equiv \Omega t$, the time dependence of the magnetization $m = N^{-1}\sum S_i$ is given by[1]

$$\Omega\frac{dm}{d\tau} = -m + \tanh\left(\frac{m + H\sin\tau}{T}\right). \tag{11.3}$$

Because of the periodic forcing, we intuitively expect that the system will settle into a forced oscillation of period 2π after the transients decay away. Our goal is to establish the characteristics of this forced oscillation.

General features

Above the phase transition, $T > T_c = 1$, the system settles into a unique forced oscillation of period 2π, independent of the initial conditions. Mathematically, Eq. (11.3) admits a

---

[1] The practically most relevant situation corresponds to $\Omega \ll 1$; accordingly, $\tau = \Omega t$ is often called the *"slow" time*.
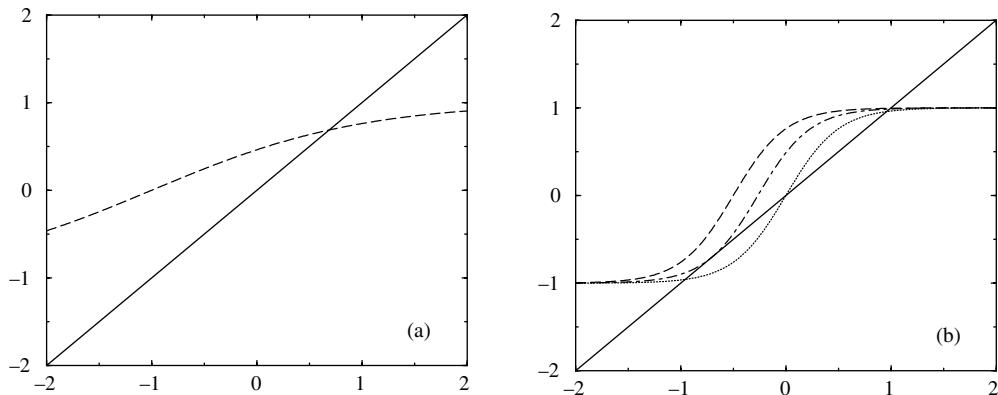
Fig. 11.2. (a) Graphical solution of Eq. (11.4) for the case of $T = 2$ and $H = 1$. The intersection of the two curves gives $m_{eq}(H = 1, T = 2)$. (b) Graphical solution of Eq. (11.4) for $T = 1/2$. The upper curve corresponds to $H = 1/2$ which exceeds $H_*(T = 1/2) \doteq 0.2664199$ (Eq. (11.5)); this curve intersects the diagonal only once. The middle and lower curves correspond to $H = H_*$ and to $H = 0$; in these cases the number of solutions is 2 and 3, respectively.

unique stable limit cycle. The proof of this assertion relies on a few basic notions about ordinary differential equations. The first step is to show that a periodic solution actually exists. We note that the equation

$$m - \tanh\left(\frac{m + H}{T}\right) = 0 \tag{11.4}$$

has a unique solution $m_{eq}(H, T)$, the equilibrium magnetization. The uniqueness of the stationary solution is evident because the slope of the curve $F(m) = \tanh[(m + H)/T]$ is smaller than 1 when $T > T_c = 1$; therefore the curve must intersect the diagonal at exactly one point (Fig. 11.2(a)).

Thus for $m > m_{eq}(H)$ the right-hand side of Eq. (11.3) is negative, so that $dm/d\tau < 0$; similarly $dm/d\tau > 0$ when $m < -m_{eq}(H)$. Hence all trajectories $m(\tau)$ eventually enter the interval $I = [-m_{eq}(H), m_{eq}(H)]$ and remain there forever. The graph of the trajectory $m(\tau)$ therefore lies in the strip $(m, \tau) \in I \times [0, \infty)$. Since $\tau = 0$ and $\tau = 2\pi$ are equivalent, we can focus on the box $I \times [0, 2\pi]$. Consider a trajectory that starts on the left side of the box and follow it until it crosses the right side. This defines a mapping of the interval I into itself, $m \to P(m)$, called the *Poincaré map*. Although it is impossible to compute the function $P(m)$ explicitly for the present example, the fixed point theorem ensures that *any* continuous[2] map of a closed interval onto itself has a fixed point, $P(m^*) = m^*$. The fixed point of the Poincaré map obviously corresponds to a forced oscillation of period 2π. Hence Eq. (11.3) has a periodic solution.

The second step is to establish uniqueness. Let us assume that two distinct periodic solutions $m_1(\tau)$ and $m_2(\tau)$ exist and obtain a contradiction. Notice that one solution lies

---

[2] Continuity follows because solutions of ordinary differential equations depend continuously on initial conditions.

strictly above the other, $m_2(\tau) > m_1(\tau)$, since trajectories cannot cross.[3] From Eq. (11.3) we obtain

$$\Omega \frac{d}{d\tau}(m_2 - m_1) = -(m_2 - m_1) + F_\tau(m_2) - F_\tau(m_1),$$

$$F_\tau(m_j) = \tanh\left(\frac{m_j + H \sin \tau}{T}\right).$$

The identity $|F_\tau(m_2) - F_\tau(m_1)| < |m_2 - m_1|$, which is valid when $T \geq T_c = 1$, implies that $m_2 - m_1$ strictly decreases. Hence $P(m_2) - P(m_1) < m_2 - m_1$ in contradiction with $P(m_j) = m_j$.

Below the phase transition, $T < T_c = 1$, there is a single *positive* stationary solution of Eq. (11.4), which we denote by $m_{eq}(H)$. Indeed, the curve $\tanh[(m + H)/T]$ lies above the origin[4] at $m = 0$ and the slope of this curve monotonically decreases with m and quickly approaches zero. Thus the curve crosses the diagonal at exactly one point where $m = m_{eq}(H)$ is positive. The crossing of the curve $\tanh[(m + H)/T]$ and the diagonal at negative m can happen only if H is sufficiently small (Fig. 11.2(b)). The threshold field H_*, beyond which the negative-m solution disappears, is called the *coercive field*. The magnitude of this coercive field is given by (see problem 11.1)

$$H_* = \sqrt{1 - T} - T \tanh^{-1}\sqrt{1 - T}. \tag{11.5}$$

For $H = H_*$, there is an additional stationary solution $-m_*$, where $m_* = \sqrt{1 - T}$; for $H < H_*$, there are two negative stationary solutions in addition to $m_{eq}(H)$.

The Poincaré map shows the existence of a periodic solution that lies in the interval $I = [-m_{eq}(H), m_{eq}(H)]$. However, there may be more than one solution. This behavior is easier to understand in the low-frequency limit. When $H < H_*$, the magnetization loop does not approach the bifurcation points $(\pm H_*, \mp m_*)$ but merely oscillates about the quasi-equilibrium curve $m_{eq}(h)$ or $-m_{eq}(h)$. These periodic solutions are called ferromagnetic since they have non-zero average magnetization. When $H > H_*$, the magnetization loop has zero average magnetization; the corresponding periodic solution is called paramagnetic. Generally for $\Omega > 0$, the phase diagram in the T–H plane contains: (i) the paramagnetic region **P** where the paramagnetic solution is stable, (ii) the ferromagnetic region **F** where two ferromagnetic solutions are stable, and (iii) the region **P** + **F** where both the paramagnetic and ferromagnetic solutions are stable. The full analytic investigation of the phase diagram is still an open problem.

11.2 Perturbation analysis

We now analyze the hysteresis loop of the mean-field ferromagnet with Glauber spin-flip dynamics, focusing on the dependence of the loop area on the amplitude H and the frequency

---

[3] Crossing is forbidden by the uniqueness of solutions of ordinary differential equations.
[4] We assume that $H > 0$; if H were negative, we could reverse its sign by shifting the time variable: $\tau \to \tau + \pi$.

Ω of the oscillating magnetic field. At the end of this section, we study hysteresis in the ferromagnetic spin chain.

High-temperature regime ($T > T_c$)

When $H = 0$, the solution to Eq. (11.3) gives the equilibrium magnetization, which vanishes in the high-temperature regime: $m_{eq}(H = 0) = 0$. Although for $H > 0$ it is impossible to represent $m_{eq}(H)$ by an explicit formula, we can treat the problem perturbatively in the $H \to 0$ limit. Expanding the hyperbolic tangent on the right-hand side of Eq. (11.3) and keeping only the linear term gives

$$\Omega \frac{dm}{d\tau} = -\left(1 - \frac{1}{T}\right) m + \frac{H}{T} \sin \tau .$$ (11.6)

This periodically forced linear equation forcing has a unique periodic solution of the form $m = M \sin(\tau - \phi)$. This expression satisfies Eq. (11.6) when the parameters M and ϕ are

$$M = \frac{H}{T} \frac{1}{\sqrt{\Omega^2 + (1 - 1/T)^2}} , \qquad \phi = \sin^{-1}\left(\frac{\Omega}{\sqrt{\Omega^2 + (1 - 1/T)^2}}\right) .$$ (11.7)

The general solution of Eq. (11.6) is the sum of this periodic solution and a homogeneous solution to Eq. (11.6). Thus

$$m = M \sin(\tau - \phi) + C e^{-(1 - T^{-1}) \tau / \Omega} ,$$ (11.8)

where the amplitude C is determined from the initial condition. Equation (11.8) shows that any solution is attracted to the periodic solution exponentially quickly, independent of the strength H of the driving field.

We now rewrite the solution $m = M \sin(\tau - \phi)$ in the form $m/M = \sin \tau \cos \phi - \cos \tau \sin \phi$ and use $h/H = \sin \tau$ to give a parametric representation of the m–h hysteresis loop. Eliminating the parameter τ yields

$$\left(\frac{h}{H}\right)^2 + \left(\frac{m}{M}\right)^2 - 2 \frac{m}{M} \frac{h}{H} \cos \phi = \sin^2 \phi .$$

By rotating the rescaled coordinates h/H and m/M by $\pi/4$,

$$\frac{h}{H} = \frac{x - y}{\sqrt{2}} , \qquad \frac{m}{M} = \frac{x + y}{\sqrt{2}} ,$$

we recast the hysteresis loop into the standard equation for an ellipse:

$$\frac{x^2}{2 \cos^2(\phi/2)} + \frac{y^2}{2 \sin^2(\phi/2)} = 1 ,$$

with semi-axes $\sqrt{2} \cos(\phi/2)$ and $\sqrt{2} \sin(\phi/2)$. Thus when the driving force is small, the dynamical hysteresis loop is an ellipse.

The angle $\vartheta = \tan^{-1}(M/H)$ between the major (horizontal) axis of the ellipse and the h axis is a decreasing function of the frequency,

$$\vartheta = \tan^{-1}\left(\frac{1}{T}\,\frac{1}{\sqrt{\Omega^2 + (1 - 1/T)^2}}\right).$$

Figure 11.1(b) schematically represents the dynamic hysteresis loop only in the low-frequency limit. In the opposite high-frequency limit, $\vartheta \simeq (T\Omega)^{-1} \to 0$, and the dynamic hysteresis loop lies along the h axis.

The area of the ellipse is $\pi \times \sqrt{2}\,\cos(\phi/2) \times \sqrt{2}\,\sin(\phi/2) = \pi \sin\phi$ in the x–y plane, so the area of the dynamical hysteresis loop in the m–h plane is $\pi MH \sin\phi$. From (11.7) we obtain

$$A = \frac{\pi}{T}\,\frac{H^2\,\Omega}{\Omega^2 + (1 - 1/T)^2}\,. \tag{11.9}$$

This expression is valid when $H \ll 1$ and agrees with the general prediction (11.2) for the exponent values $a = \alpha = 2$ and $b = \beta = 1$.

The above asymptotic results hold when the magnetic field is weak. We now show that they remain true in general. For example, in the mean-field Ising model, perturbation techniques show that the area of the dynamical hysteresis loop continues to exhibit the limiting behaviors $A \sim \Omega$ for $\Omega \ll 1$ and $A \sim \Omega^{-1}$ for $\Omega \gg 1$, independent of the driving field strength.

Example 11.1. *Low-frequency limit.* We seek a solution to Eq. (11.3) as the power series

$$m(\tau) = m_0(\tau) + \Omega\, m_1(\tau) + \Omega^2\, m_2(\tau) + \cdots\,. \tag{11.10}$$

Substituting (11.10) into (11.3) we find recursive equations for $m_j(\tau)$. At zeroth order in Ω, we obtain

$$m_0(\tau) = \tanh\left(\frac{m_0(\tau) + H\sin\tau}{T}\right), \tag{11.11}$$

while the first-order correction is

$$m_1(\tau) = -\frac{dm_0}{d\tau}\left(1 - \frac{1 - m_0^2(\tau)}{T}\right)^{-1}. \tag{11.12}$$

Similarly one can determine higher-order corrections (problem 11.2). These terms have a limited value, however, because the area of the hysteresis loop is generally proportional to Ω (see (11.14) below). Hence the first-order term actually provides the leading contribution.

A formal solution to Eqs (11.11) and (11.12) is

$$m_0(\tau) = m_{\rm eq}(h), \qquad m_1(\tau) = -\frac{dh}{d\tau}\frac{dm_{\rm eq}}{dh}\left(1 - \frac{1 - m_{\rm eq}^2(h)}{T}\right)^{-1}. \tag{11.13}$$

That is, the zeroth-order contribution $m_0(\tau)$ coincides with the equilibrium solution. The area of the dynamical hysteresis loop $A = \oint m \, dh$ vanishes in zeroth order, since there is no time lag between $m_0(\tau) = m_{eq}(h)$ and h. Since $m_1(\tau) \sim dh/d\tau$, the time lag between $m_1(\tau)$ and $h(\tau)$ equals $\pi/2$, a quarter of the period. Hence the leading contribution to the area of the dynamical hysteresis loop is

$$A = \Omega \int_0^{2\pi} m_1(\tau) \, \frac{dh}{d\tau} \, d\tau \, . \tag{11.14}$$

Example 11.2. *High-frequency limit.* When $\Omega \gg 1$, we use Ω^{-1} as the expansion parameter and write[5]

$$m(\tau) = m_0(\tau) + \Omega^{-1} \, m_1(\tau) + \Omega^{-2} \, m_2(\tau) + \cdots \, . \tag{11.15}$$

Substituting (11.15) into (11.3) we find

$$\frac{dm_0}{d\tau} = 0, \qquad \frac{dm_1}{d\tau} = -m_0(\tau) + \tanh \left(\frac{m_0(\tau) + H \sin \tau}{T} \right),$$

etc. The first equation shows that $m_0(\tau)$ is a constant; the next equation admits a periodic solution only when $m_0(\tau) = 0$. Therefore the leading behavior of the magnetization is

$$m(\tau) = \Omega^{-1} \, m_1(\tau), \qquad m_1(\tau) = \int_0^\tau \tanh \left(\frac{H}{T} \, \sin \xi \right) d\xi + C_1 \, , \tag{11.16}$$

where the constant C_1 is fixed by the requirement that the next term $m_2(\tau)$ of the expansion (11.15) is a periodic function of τ. From $m \sim \Omega^{-1}$ and $h \sim 1$ we conclude again that the dynamic hysteresis loop lies along the h axis in the high-frequency limit. The leading contribution to the area of the dynamical hysteresis loop is

$$A = \Omega^{-1} \int_0^{2\pi} m_1(\tau) \, \frac{dh}{d\tau} \, d\tau \, . \tag{11.17}$$

Substituting $m_1(\tau)$ from (11.16) into (11.17) and integrating by parts, we obtain

$$A = \frac{4H}{\Omega} \int_0^1 \tanh \left[\frac{H}{T} \, \sqrt{1 - x^2} \right] dx \, . \tag{11.18}$$

The previous result (11.9) $A = \pi H^2 / \Omega T$ is recovered when $H \ll T$.

Critical regime ($T = T_c$)

In the critical regime, the system still settles into a unique forced oscillation whose qualitative behavior is the same as that in the high-temperature regime, but quantitative differences arise. For instance, when the external field is small, the term linear in m on the right-hand side

---

[5] We use the same notation $m_j(\tau)$ as in Eq. (11.10) although the functions are different.

of (11.6) vanishes; thus Eq. (11.6) does not provide a proper approximation to Eq. (11.3). Instead, we must also keep the next (cubic) term in the expansion to obtain a consistent approximation.

Let us now consider the high- and low-frequency limits in detail. In the former case, the nature of hysteresis is the same as in the high-temperature regime. The results (11.15)–(11.18) continue to hold and the area of the dynamical hysteresis loop is given by Eq. (11.18), with T set to 1:

$$A = \frac{4H}{\Omega} \int_0^1 \tanh\left[H\sqrt{1-x^2}\right] dx. \tag{11.19}$$

The situation is richer in the low-frequency limit. Let us again seek a series solution of the form (11.10). This gives the solutions (11.11) and (11.12) with $T = 1$, namely

$$m_0 = m_{\text{eq}} = \tanh(m_0 + h), \qquad m_1(\tau) = \frac{d}{d\tau}\left(\frac{1}{m_0}\right). \tag{11.20}$$

We now additionally assume that the external field is small, so that m_0 is also small. Expanding the hyperbolic tangent in the first of Eqs (11.20) gives

$$m_0 = m_0 + h - \frac{1}{3}(m_0 + h)^3 + \cdots,$$

from which $m_0 \simeq (3h)^{1/3}$ as $h \to 0$. Thus $m_0 \simeq (3H\tau)^{1/3}$ near $\tau = 0$. The second of Eqs (11.20) then yields $m_1 \simeq -H^{-1/3}(3\tau)^{-4/3}$. This singularity shows that the expansion (11.10), which assumes $m_0 \gg \Omega m_1$, breaks down when $m_0 \sim \Omega m_1$, i.e. for $\tau \sim H^{-2/5}\Omega^{3/5}$. This kind of inconsistency often arises in applying a naive perturbation expansion. The appropriate way to overcome this difficulty is to apply the boundary layer technique. The essence of this technique can be summarized as follows:

1. The series (11.10) is regarded as the *outer expansion* which is valid in the outer region, which is defined by $\tau \gg \Omega^{3/5}$ in the present case.
2. The region complementary to the inner region is usually termed the *boundary layer*. Within the boundary layer we use the properly rescaled variables and seek a solution as an *inner expansion* in that variable.
3. The inner and outer solutions are matched in the overlap region.

Since the width of the boundary layer is proportional to $H^{-2/5}\Omega^{3/5}$, we begin the boundary layer method method by changing from the basic variable τ to θ by the transformation

$$\tau = \Omega^{3/5} H^{-2/5} \theta. \tag{11.21}$$

When $\theta \gg 1$, we are still in the outer region, so that $m \simeq (3H\tau)^{1/3} = (\Omega H)^{1/5}(3\theta)^{1/3}$. This form suggests that the correct scaling of the dominant term in the inner expansion is given by

$$m(\tau) = (\Omega H)^{1/5} M_0(\theta) + \cdots. \tag{11.22}$$

Substituting (11.21)–(11.22) into (11.3) we obtain

$$\Omega \frac{\Omega^{1/5} H^{1/5}}{\Omega^{3/5} H^{-2/5}} \frac{dM_0}{d\theta} = -(\Omega H)^{1/5} M_0 + \tanh\left[(\Omega H)^{1/5} M_0 + (\Omega H)^{3/5}\theta\right].$$

Expanding the hyperbolic tangent and equating the leading-order terms gives

$$\frac{dM_0}{d\theta} = \theta - \frac{1}{3} M_0^3. \tag{11.23}$$

The dominant remaining term is $\Omega H \times (-\theta M_0^2)$. To cancel it we introduce the correction $(\Omega H)^{3/5} M_1(\theta)$. Comparing the scale of the correction with the scale of the leading term (11.22), we conclude that the appropriate perturbation parameter is $(\Omega H)^{2/5}$. Therefore the inner expansion becomes

$$m(\tau) = (\Omega H)^{1/5} \left[M_0(\theta) + (\Omega H)^{2/5} M_1(\theta) + (\Omega H)^{4/5} M_2(\theta) + \cdots\right]. \tag{11.24}$$

Substituting (11.24) into (11.3), we recover (11.23) to leading order and we obtain

$$\frac{dM_1}{d\theta} = -[\theta + M_1]M_0^2$$

at the next order.

The inner and outer expansions are valid when $|\tau| \ll 1$ and $|\theta| \gg 1$, respectively, and we now match the expansions in the overlap region $\Omega^{3/5} H^{-2/5} \ll \tau \ll 1$.[6] This matching is obvious to leading order – requiring $M_0 \simeq (3\theta)^{1/3}$ when $|\theta| \to \infty$, we recover the asymptotic $m \simeq (3H\tau)^{1/3}$, which applies when $|\tau| \ll 1$. The matching of higher-order corrections is significantly more involved. Fortunately, we do not need these higher-order corrections, since the leading term $\int m_0\, dh$ does not vanish. Finally, we use the relation $m_0\, dh = (\Omega H)^{1/5} M_0(\theta) \times (\Omega H)^{3/5}\, d\theta = (\Omega H)^{4/5} M_0(\theta)\, d\theta$, and see that the dominant contribution to the area of the dynamical hysteresis loop is obtained in the boundary layer. Since the boundary layer around $\tau = \pi$ gives the same contribution as the layer $\tau = 0$, we obtain

$$A = C\,\Omega^{4/5} H^{4/5}, \qquad C = -2\int_{-\infty}^{\infty} M_0(\theta)\, d\theta, \tag{11.25}$$

for the area of the dynamical hysteresis loop in the low-frequency limit.[7]

Thus the area of the dynamical hysteresis loop is given by (11.19) and (11.25) in the high- and low-frequency limits, respectively. The latter result is valid when additionally $H \ll 1$. Making this same small-H approximation in Eq. (11.19), the asymptotic results are

$$A(T_c, H, \Omega) \simeq \begin{cases} \pi H^2\, \Omega^{-1}, & H^{2/3} \ll \Omega,\ H \ll 1, \\ C\, H^{4/5}\, \Omega^{4/5}, & \Omega \ll H^{2/3} \ll 1. \end{cases} \tag{11.26}$$

---

[6] When the driving magnetic field is small, the proper definition of the low-frequency limit is $\Omega \ll H^{2/3}$ rather than $\Omega \ll 1$.

[7] The constant C is indeed positive (problem 11.4).

These limiting behaviors suggest that the loop area satisfies a dynamical scaling ansatz of the form

$$A(T_c, H, \Omega) = H^{4/5} \, \Omega^{4/5} \, \mathcal{A}(z), \qquad z = \frac{\Omega}{H^{2/3}} \, ,$$

which should hold in the scaling limit $H \to 0$ and $\Omega \to 0$, with z kept fixed. Equation (11.26) then gives $\mathcal{A}(0) = C$ and $\mathcal{A}(z) \to \pi \, z^{-9/5}$ when $z \gg 1$.

Low-temperature regime ($T < T_c$)

Consider first the most interesting low-frequency limit. For $H < H_*$, periodic solutions are ferromagnetic. Seeking $m(\tau)$ as a series of the form (11.10), we find (11.11) and (11.12) for m_0 and m_1. The area of the dynamical hysteresis loop is linear in Ω and is given by (11.14). The leading contribution to the average magnetization is

$$M = \frac{1}{2\pi} \int_0^{2\pi} m_0(\tau) \, d\tau = \frac{1}{\pi} \int_{-H}^{H} \frac{m_{\mathrm{eq}}(h)}{\sqrt{H^2 - h^2}} \, dh \, .$$

If we additionally assume that the magnetic field is small, $h \ll 1$, the average magnetization would be close to $\mu \equiv m_{\mathrm{eq}}(H = 0)$.[8] To simplify the expression (11.14) for the loop area, we take the Taylor expansion for the equilibrium magnetization,

$$m_{\mathrm{eq}}(h) = m_{\mathrm{eq}}(0) + \frac{dm_{\mathrm{eq}}}{dh}\bigg|_{h=0} h + \cdots = \mu + \left(\frac{T}{1 - \mu^2} - 1 \right)^{-1} h + \cdots \, ,$$

and substitute it into (11.12) to yield

$$m_1 = - \left(\frac{T}{1 - \mu^2} - 1 \right)^{-1} \left(1 - \frac{1}{T} \frac{1}{1 - \mu^2} \right)^{-1} \frac{dh}{d\tau}$$

to lowest order. The area of the dynamical hysteresis loop becomes

$$A = \pi H^2 \Omega \, \frac{(1 - \mu^2)^2}{1 + (1 - \mu^2)^2 - (1 - \mu^2)(T^{-1} + T)} \, , \tag{11.27}$$

when H and Ω are small. From Eq. (11.27), the dependence of the area on H and frequency Ω is similar to that in Eq. (11.9) that applies above the critical temperature, while the temperature dependence is more complicated:

$$A(T, H, \Omega) \to \pi H^2 \Omega \times \begin{cases} 16 \cdot e^{-4/T}, & T \to 0, \\ \frac{1}{8} (1 - T)^{-2}, & T \to 1 \, . \end{cases} \tag{11.28}$$

For $H > H_*$, the periodic solution is paramagnetic. In the low-frequency limit, the area of the dynamical hysteresis loop approaches the area of the static loop:

$$A(T, H, \Omega) \to A_0 = 2 \int_{-H_*}^{H_*} m_{\mathrm{eq}}(h) \, dh \, .$$

---

[8] The quantity μ is the positive solution to $\mu = \tanh(\mu/T)$.

The extra area beyond A_0 scales as $\Omega^{2/3}$. There is no elementary derivation of this scaling, and here we merely describe the idea. The derivation is technically easier if one assumes that the temperature is close to T_c. Next we expand the hyperbolic tangent on the right-hand side of Eq. (11.3) to give the time-dependent Ginzburg–Landau equation with a periodic driving force:

$$\Omega \frac{dm}{d\tau} = (1 - T)m - \frac{1}{3} m^3 + H \sin \tau .$$

The static hysteresis loop follows the unstable branch until $(m, h) = (m_*, -H_*)$ with $m_* = \sqrt{1 - T}$ and $H_* = \frac{2}{3} (1 - T)^{3/2}$, then jumps to $(m, h) = (-3\sqrt{1 - T}, -H_*)$, and follows the stable branch. If we seek a solution as the regular expansion (11.10), we would find a singularity near $h = H_*$. This singularity is resolved by introducing the boundary layer region. The analysis is similar to that described above for the critical regime and gives the scales of the distortion of the dynamical hysteresis loop from the static hysteresis loop. If both the frequency is small and the amplitude H slightly exceeds the coercive field H_*, the extra area scales as

$$A(T, H, \Omega) - A_0 \sim \Omega^{2/3} (H - H_*)^{1/3}.$$

Hysteresis for the ferromagnetic spin chain

While the one-dimensional ferromagnetic Ising spin chain is somewhat pathological, since $T_c = 0$ and therefore there is no static hysteresis, this system is especially simple and therefore well suited to understand dynamic hysteresis for $T > 0$. Using the standard Glauber spin-flip rate

$$w_i = \frac{1}{2} \left(1 - s_i \tanh \left[\frac{J(s_{i-1} + s_{i+1}) + h(t)}{T} \right] \right),$$

the local magnetization $S_i = \langle s_i \rangle$ evolves according to

$$\frac{dS_i}{dt} = -S_i + \left\langle \tanh \left[\frac{J(s_{i-1} + s_{i+1}) + h(t)}{T} \right] \right\rangle.$$

As we learned in Chapter 8, it is straightforward to perform the average when $h = 0$:

$$\left\langle \tanh \left[\frac{J(s_{i-1} + s_{i+1})}{T} \right] \right\rangle = \left\langle \gamma \frac{s_{i-1} + s_{i+1}}{2} \right\rangle = \frac{\gamma}{2} (S_{i-1} + S_{i+1})$$

where $\gamma = \tanh(2J/T)$. This trick does not work when $h \neq 0$, and it is not possible to derive closed equations for the local magnetization in this case. However, analytic progress is possible for weak fields $H \ll \min(J, T)$. Now we perform the expansion

$$\tanh \left[\frac{J(s_{i-1} + s_{i+1}) + h}{T} \right] = \gamma \frac{s_{i-1} + s_{i+1}}{2} + \frac{h}{T} \left[1 - \gamma^2 \left(\frac{s_{i-1} + s_{i+1}}{2} \right)^2 \right] + \cdots ,$$

and then average to obtain

$$\frac{dS_i}{dt} = -S_i + \frac{\gamma}{2}(S_{i-1} + S_{i+1}) + \frac{h}{T}\left[1 - \gamma^2 \frac{1 + \langle s_{i-1}s_{i+1}\rangle}{2}\right] + \cdots.$$

Summing these equations gives the closed equation for the magnetization $m = N^{-1}\sum S_i$,

$$\frac{dm}{dt} = -(1 - \gamma)m + \frac{h}{T}\left[1 - \frac{1}{2}\gamma^2(1 + G_2)\right] + \cdots, \qquad (11.29)$$

where $G_k = \langle s_i s_{i+k}\rangle$ is the spin correlation function. After the transients decay away, the magnetization settles into a forced periodic oscillation with $m \sim h$. The spin correlation function also settles into a periodic nutation of the order of h around the equilibrium value $G_k^{\mathrm{eq}} = \eta^k$, with $\eta = \tanh(J/T)$. Since G_2 is multiplied by $h \ll 1$, we ignore the contribution due to nutation.[9] Replacing G_2 by η^2 and keeping only terms up to order $\mathcal{O}(h)$ we convert (11.29) into

$$\Omega\frac{dm}{d\tau} = -(1 - \gamma)m + \frac{H}{T}\left[1 - \frac{\gamma^2(1 + \eta^2)}{2}\right]\sin\tau.$$

Apart from the coefficients, this equation is identical to (11.6). Therefore the counterparts of the subsequent results (11.7)–(11.9) hold. In particular, the dynamical hysteresis loop is an ellipse of area[10]

$$A = \frac{\pi}{T}\frac{1 - \eta^2}{1 + \eta^2}\frac{H^2\Omega}{\Omega^2 + (1 - \gamma)^2},$$

and the amplitude of the magnetization is

$$M = \frac{H}{T}\frac{1 - \eta^2}{1 + \eta^2}\frac{1}{\sqrt{\Omega^2 + (1 - \gamma)^2}}. \qquad (11.30)$$

In the low-frequency limit, this expression reduces to the equilibrium static susceptibility of a ferromagnetic Ising chain (problem 11.5).

11.3 Disordered ferromagnets

In most applications, materials that exhibit hysteresis are at temperatures substantially smaller than the critical temperature, so it is useful to consider the case of zero temperature. In the adiabatic limit ($\Omega \to 0$), the hysteresis loop of homogeneous ferromagnets at zero temperature is rectangular when $H > H_*$.[11] However, real hysteresis loops have different

---

[9] If we were to keep it, we should have expanded the hyperbolic tangent to order $\mathcal{O}(h^2)$.

[10] We have simplified various formulae by using the identity $\gamma = 2\eta/(1 + \eta^2)$.

[11] This result holds for arbitrary spatial dimension; only the magnitude H_* of the coercive field depends on the dimension.

shapes and they exhibit magnetization jumps as the field is slowly varied. One might expect that these jumps would disappear in careful measurements; nevertheless, they persist. Perhaps the most impressive manifestation of this jerky response is *Barkhausen noise*, which is produced by the sudden reorientations of local magnetizations as the field varies.

The simplest theoretical framework that at least qualitatively accounts for some of the experimental behavior is the *random-field Ising model* (RFIM), where each spin experiences a distinct magnetic field.[12] The Hamiltonian of the RFIM is

$$\mathcal{H} = -J \sum_{\langle ij \rangle} s_i s_j - \sum_i h_i s_i - h \sum_i s_i \,,$$

where the on-site magnetic fields h_i are quenched, independent, and identically distributed random variables that are chosen from the probability density $\rho(h_i)$. We assume that $\rho(h_i)$ is continuous and set $\overline{h_i} = 0$ at the outset.[13] If the disorder distribution is even, $\rho(h_i) = \rho(-h_i)$, then additionally the hysteresis loop is symmetric, i.e. invariant under $m \to -m$ and $h \to -h$. A crucial assumption for the following analysis is that the spins interact ferromagnetically ($J > 0$). We focus on static hysteresis, where the external magnetic field h changes adiabatically unless stated otherwise; equivalently, all flippable spins relax instantly. At zero temperature, only energy-lowering flips are allowed, and each spin s_i aligns with the local field L_i at site i,

$$L_i \equiv J \sum_{j \in \langle i \rangle} s_j + h_i + h \,,$$

namely, the sum of the external homogeneous and random fields, as well as the interactions due to the neighboring spins s_j of spin s_i. At zero temperature $s_i = \mathrm{sgn}(L_i)$.

Before discussing the dynamics of the RFIM, we show that the random field can drastically affect ferromagnetic order at zero temperature. Starting with the ferromagnetic state where all spins point up, consider the energy change of a compact droplet of radius R that flips to pointing down. There are two contributions to the energy change of this droplet: (i) an increase of $J R^{d-1}$ due to the creation of the droplet interface; and (ii) a decrease of $\delta h R^{d/2}$, where δh is the typical value of the random field, because the spins align with the average of the random field in a region of volume R^d. The energy cost of this droplet is positive if $J R^{d-1} > \delta h R^{d/2}$, or $d > 2$. Thus Ising ferromagnetism at zero temperature is stable to random-field perturbations above two spatial dimensions.

We now describe general properties of the RFIM that are independent of the spatial dimension and the lattice structure. Let's start with $h = \infty$, where all spins have the value $+1$. As h decreases, eventually the local field L_i at some site i first crosses zero. At this point, the ith spin flips: $s_i = -1$. This event reduces the local field of neighboring

---

[12] The RFIM has an advantage of being more tractable than more exotic magnetic models, such as the spin glass. Some qualitative features of the RFIM are robust and therefore its analysis sheds light on hysteresis in other disordered ferromagnets.

[13] Here $\overline{(\ldots)}$ denotes the average of a quantity $(\ldots)$ over the disorder. If the average of the random field is non-zero, we can make it zero by subtracting the average from the random field at each site and adding this average to the external field.
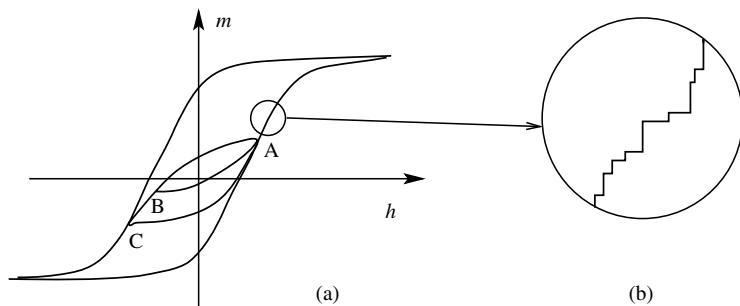
(a) Static hysteresis for the RFIM. (b) Magnification of the hysteresis curve to illustrate magnetization jumps that give Barkhausen noise. Also shown are minor (return) loops that lie inside the main loop; these are obtained by cycling the field over a reduced range of h.

spins, possibly causing some of them to flip. In turn, the local fields of additional spins are diminished, possibly leading to addition spin flips, etc. Hence a single spin flip can create a concomitant avalanche of spin flips. Once the avalanche terminates, we then decrease the external magnetic field h again until another spin flips, possibly causing another avalanche. The process continues until all spins flip, and the outcome is the upper half of the hysteresis loop $m_+(h)$.

The magnetization remains constant between avalanches, thereby giving the jumps in the hysteresis loop and the corresponding Barkhausen noise (Fig. 11.3). Another interesting feature of the RFIM is the existence of minor return loops. Such a loop is obtained, for example, by starting at point A of the lower half of the main loop and decreasing the field until the upper half of the minor loop crosses the upper half of the main loop (point C). If the field is then increased, the lower part of the minor loop arises which returns exactly to the starting point A. If the decrease of the field was stopped a bit earlier (at point B, for example), another minor loop would be obtained, which also returns exactly to the starting point A. While these same features arise for *any* disordered system with zero-temperature spin-flip dynamics, certain properties are specific to the ferromagnetic RFIM. We list them here because of their utility in analyzing hysteresis loops.

1. *Abelian property*. The order in which eligible spins (those anti-aligned with the local field) are flipped during an avalanche does not matter.
2. *One flip*. As h changes from $+\infty$ to $-\infty$, each spin flips once and only once.
3. *Symmetry*. The lower half of the hysteresis loop $m_-(h)$ that is obtained by increasing the magnetic field from $h = -\infty$ to $h = +\infty$ is the mirror image of the upper half:[14] $m_-(h) = -m_+(-h)$.
4. *No passing*. Denote $\mathbf{s} = \{s_1, \ldots, s_N\}$ as the entire spin configuration. Suppose two configurations $\mathbf{s}^{(1)}(t)$ and $\mathbf{s}^{(2)}(t)$ evolve due to the influence of the fields $h^{(1)}(t)$ and

---

[14] This condition applies only in the thermodynamic limit when the random-field distribution is even: $\rho(h_i) = \rho(-h_i)$.

$h^{(2)}(t)$ that satisfy $h^{(1)}(t) \geq h^{(2)}(t)$. If one configuration initially "exceeds" the other, that is, $\mathbf{s}^{(1)}(0) \geq \mathbf{s}^{(2)}(0)$ (a shorthand for $s_i^{(1)}(0) \geq s_i^{(2)}(0)$ for all $i = 1, \ldots, N$), then $\mathbf{s}^{(1)}(t) \geq \mathbf{s}^{(2)}(t)$ for all t.

5. *Return-point memory*. Suppose $\mathbf{s}(0)$ evolves due to the influence of the field $h(t)$ that satisfies $h(0) \leq h(t) \leq h(t_f)$. The final state $\mathbf{s}(t_f)$ depends only on $h(t_f)$; it does *not* depend on the duration t_f and the history $h(t)$.

To understand the Abelian property, suppose that a flip of a given spin i has made two of its neighbors at j and k flippable (i.e. L_j and L_k become negative). If s_j flips first, this would further decrease the local field at k, if the site k is a neighbor of j (when all couplings are ferromagnetic); otherwise, the local field at k would not change. In either case, the spin at k remains flippable. Similarly flipping s_k first would leave s_j flippable. For any other site $\ell \neq j, k$, either order of flipping s_j and s_k leads to the same resulting local field at site ℓ. Repeating this argument for subsequent flips, we conclude that the order of flipping is indeed irrelevant for the final outcome and the entire avalanche will involve the same spins. The second and third properties are obvious. To justify the no-passing rule, suppose that passing was possible and it first occurred at site i, i.e. $s_i^{(1)}(0) > s_i^{(2)}(0)$, but $s_i^{(1)}(t) < s_i^{(2)}(t)$. Then the local fields would satisfy $L_j^{(1)}(t) < 0 < L_j^{(2)}(t)$. But this is impossible since $s_j^{(1)}(t) \geq s_j^{(2)}(t)$ and $h^{(1)}(t) \geq h^{(2)}(t)$. Finally, the return-point memory is clearly valid if the field monotonically increases (or decreases) with time. To handle the general situation when $h(t)$ is not necessarily monotonic, we construct two auxiliary fields $h_{\min}$ and $h_{\max}$ defined as

$$h_{\min}(t) = \min_{t': t \leq t' \leq t_f} h(t'), \qquad h_{\max}(t) = \max_{t': 0 \leq t' \leq t} h(t').$$

These auxiliary fields are monotonic by construction and the original field is sandwiched between them: $h_{\min}(t) \leq h(t) \leq h_{\max}(t)$. Let $\mathbf{s}_{\min}(t)$, $\mathbf{s}(t)$, and $\mathbf{s}_{\max}(t)$, all starting from $\mathbf{s}(0)$, evolve under $h_{\min}(t)$, $h(t)$, and $h_{\max}(t)$. The no-passing rule tells us that $\mathbf{s}_{\min}(t) \leq \mathbf{s}(t) \leq \mathbf{s}_{\max}(t)$. We know that $\mathbf{s}_{\min}(t_f) = \mathbf{s}_{\max}(t_f)$ due to the already established return-point memory property for the monotonic paths.[15] Thus $\mathbf{s}(t_f)$ must also depend only on t_f for an arbitrary (not necessarily monotonic) $h(t)$.

The properties outlined above underlie static hysteresis in the ferromagnetic RFIM. More subtle characteristics, such as the avalanche size distribution, depend on the precise form of the distribution of random fields. A more serious obstacle to theoretical progress is the spatial dimensionality. Generally, there are no techniques that theoretically handle hysteresis in the ferromagnetic RFIM on lattices with loops; consequently, hysteresis in the two- and three-dimensional RFIM has been primarily investigated numerically. However, hysteresis is tractable on a one-dimensional (loopless) lattice and also on the complete graph that has a maximal number of loops. We now discuss the latter system (often called the mean-field model) because of the valuable insights that this example provides.

---

[15] The field is assumed to remain within the bounds, $h(0) \leq h(t) \leq h(t_f)$. Therefore the definition of auxiliary fields $h_{\min}$ and $h_{\max}$ implies $h_{\min}(0) = h(0) = h_{\max}(0)$ and $h_{\min}(t_f) = h(t_f) = h_{\max}(t_f)$.

11.4 Mean-field model

For the mean-field, or complete graph, model, all pairs of sites are connected. Setting the coupling constant to $1/N$, the local field at site i is $m + h_i + h$. Spins that point down are those where the random magnetic field obeys $h_i < -m - h$. The fraction of such spins is $\int_{-\infty}^{-m-h} \rho(x)\, dx$, so that the magnetization is given by

$$m(h) = 1 - 2 \int_{-\infty}^{-m(h)-h} \rho(x)\, dx. \tag{11.31}$$

Equation (11.31) is generally unsolvable, but some simple disorder distributions are tractable (Fig. 11.4).

Example 11.3. *Uniform disorder distribution.* For the uniform distribution of random magnetic fields,

$$\rho(h_i) = (2\Delta)^{-1} \times \begin{cases} 1, & |h_i| \le \Delta, \\ 0, & |h_i| > \Delta, \end{cases} \tag{11.32}$$

we find, from (11.31), that $m(h) = 1$ if $-m(h) - h < -\Delta$, or equivalently $h > \Delta - 1$. This result indicates that the critical disorder strength is $\Delta_c = 1$, below which a rectangular hysteresis loop arises (Fig. 11.4(a)). Continuing this analysis (problem 11.6), when $\Delta > \Delta_c = 1$, the magnetization is given by

$$m(h) = \begin{cases} -1, & h \le -(\Delta - 1), \\ h/(\Delta - 1), & |h| < \Delta - 1, \\ +1, & h \ge +(\Delta - 1). \end{cases} \tag{11.33}$$

(a)   (b)
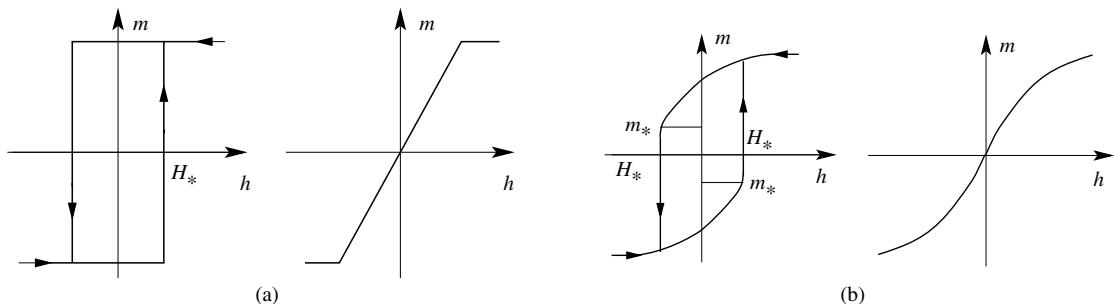
Fig. 11.4. Hysteresis loops and magnetization curves. (a) Uniform distribution of random fields. For weak disorder ($\Delta < 1$, left side), the coercive field is $H_* = 1 - \Delta$, while the hysteresis loop disappears for strong disorder ($\Delta \ge 1$, right side). (b) Exponential random-field distribution. For weak disorder ($\Delta < 1$, left side)), infinite avalanches occur at the bifurcation points ($m_*, -H_*$) and ($-m_*, H_*$), with $m_* = 1 - \Delta$ and $H_* = 1 - \Delta + \Delta \ln \Delta$, while the hysteresis loop disappears for strong disorder ($\Delta \ge 1$, right side).

The disappearance of the hysteresis loop for sufficiently weak disorder is generic, while its shape depends on the disorder distribution; a rectangular loop is specific to the uniform disorder distribution.

Example 11.4. *Exponential disorder distribution*. For the disorder distribution

$$\rho(h_i) = (2\Delta)^{-1} e^{-|h_i|/\Delta}, \tag{11.34}$$

the hysteresis loop also disappears when the disorder exceeds the critical strength $\Delta_c = 1$ (Fig. 11.4(b)). To understand the emergence of this critical strength we differentiate Eq. (11.31) to eliminate the integral. We find

$$\frac{dm}{dh} = \frac{2\rho(-m-h)}{1 - 2\rho(-m-h)} = \frac{e^{-|m+h|/\Delta}}{\Delta - e^{-|m+h|/\Delta}}, \tag{11.35}$$

where the last equality applies to the exponential distribution (11.34).

When $\Delta \geq 1$, the denominator in Eq. (11.35) remains positive and m smoothly increases from $m(-\infty) = -1$ to $m(\infty) = 1$ so that there is no hysteresis loop. When $\Delta < \Delta_c = 1$, the denominator vanishes, and the magnetization jumps to another branch.

We substitute the exponential distribution (11.34) into (11.31) to obtain the magnetization

$$m = \begin{cases} e^{(m+h)/\Delta} - 1, & m+h < 0, \\ 1 - e^{-(m+h)/\Delta}, & m+h > 0. \end{cases} \tag{11.36}$$

When $\Delta \geq 1$, we expand Eq. (11.36) near the origin to give (problem 11.6)

$$m = \begin{cases} \sqrt{2h} - h + \cdots, & \Delta = 1, \\ \dfrac{1}{\Delta - 1} h - \dfrac{1}{2} \dfrac{\Delta}{(\Delta - 1)^3} h^2 + \cdots, & \Delta > 1. \end{cases} \tag{11.37}$$

Suppose now that $\Delta < \Delta_c = 1$. If we decrease the magnetic field, the magnetization will jump from the top branch to the lower branch when $(m, h) = (m_*, -H_*)$ is reached. If we then increase the magnetic field, the jump back to the top branch will occur at $(m, h) = (-m_*, H_*)$. The first line in Eq. (11.36) represents this top branch. Therefore at the bifurcation point $(m_*, -H_*)$ we get $m_* = 1 - e^{-(m_*-H_*)/\Delta}$. At this point the denominator in Eq. (11.35) must vanish. This gives $\Delta = e^{-(m_*-H_*)/\Delta}$. From these two relations we find $m_* = 1 - \Delta$ and $H_* = 1 - \Delta + \Delta \ln \Delta$.

Generally, for an arbitrary symmetric distribution of random magnetic fields that has a single maximum at the origin, that is, $\rho(-x) \equiv \rho(x)$ and $\rho(x) \leq \rho(0)$ for all x, we use Eq. (11.35) and find that the relation $\rho(0) = 1/2$ defines the case of critical disorder. When the disorder strength exceeds the critical value, the magnetization scales linearly with h,

$$m = \frac{h}{\Delta - 1} + \mathcal{O}(h^2), \tag{11.38}$$

in the small-field limit (problem 11.6). To determine the small-h behavior for critical disorder strength, we must know the behavior of the random-field distribution near the origin. If $\rho(h_i) = \frac{1}{2} - A|h_i|^\mu + \cdots$ when $|h_i| \ll 1$, then

$$m \simeq \left(\frac{1+\mu}{2A} \right)^{1/(1+\mu)} h^{1/(1+\mu)} \tag{11.39}$$

for $0 < h \ll 1$. This non-analytic behavior is non-universal as it depends on the details (particularly the exponent μ) of the random-field distribution.

Avalanche size distribution

Avalanches are the salient feature of hysteresis in disordered ferromagnets when the magnetic field is adiabatically varied. Here we study their size distribution. Suppose that an avalanche starts at (m, h); i.e. the most susceptible spin with random field $x = -m - h$ flips down. This spin-flip event reduces the magnetization from m to $m - 2/N$ (see Fig. 11.5). The avalanche will terminate if there are no other spins whose random fields are in the range $[x, x + 2/N]$. The average number of such spins is $\lambda \equiv 2\rho(-m - h)$, so that the probability that no spin flips is $e^{-\lambda}$. Hence the avalanche size is equal to 1 with probability $A_1(\lambda) = e^{-\lambda}$. Similarly the first spin flip causes just one more spin flip when there is exactly one spin with random field in the range $[x, x + 2/N]$ and no other spin with random field in the range $[x + 2/N, x + 4/N]$. The probability of this event is $A_2(\lambda) = \lambda e^{-\lambda} \times e^{-\lambda} = \lambda e^{-2\lambda}$. Continuing this line of reasoning (problem 11.7), the avalanche size distribution is

$$A_s(\lambda) = \frac{(s\lambda)^{s-1}}{s!} e^{-s\lambda} . \tag{11.40}$$
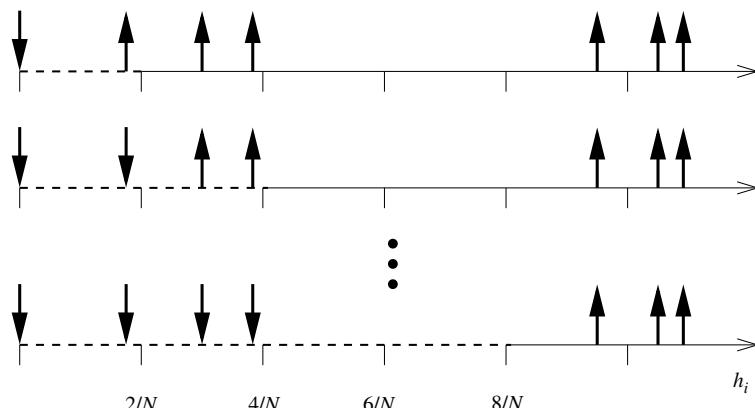
Fig. 11.5. Illustration of an avalanche in the mean-field RFIM. The spins are arranged in increasing order of their random magnetic fields. After the most susceptible spin flips, a spin whose random field h_i lies within the dashed segment (within $2/N$ of the random field of the most susceptible spin) flips. As shown, an avalanche (of size 4) terminates when all up spins have their random fields h_k outside the dashed segment after four spins have flipped.

We have already encountered this nice formula in the context of product-kernel aggregation (see Eq. (5.44)). In this system, the cluster size distribution $c_s(t)$ is given by $s^{-1}A_s(t)$! The statement of mass conservation, $\sum_{s\geq 1} sc_s(t) = 1$, therefore implies the normalization $\sum_{s\geq 1} A_s(\lambda) = 1$.

The above results are general, but the dependence on the random-field distribution is hidden in the parameter $\lambda = 2\rho(-m-h)$. Let us now consider two concrete examples to determine the nature of the avalanche size distribution in the m–h plane.

Example 11.5. *Uniform disorder distribution*. In this case we have $-m-h < -1+H_* = -\Delta$ on the upper branch. Since $\rho(x) = 0$ when $x < -\Delta$ for the uniform disorder distribution (11.32), we get $\lambda = 2\rho(-m-h) = 0$. Thus in agreement with Fig. 11.4(b), there are no avalanches, except for the infinite avalanche that brings the system from $(m, h) = (1, -H_*)$ all the way down to $(m, h) = (-1, -H_*)$.

Example 11.6. *Exponential disorder distribution*. Now $\lambda > 0$ and the avalanche size distribution is non-trivial everywhere. Along the upper branch defined by $m = 1 - e^{-(m+h)/\Delta}$, the point (m, h) flows from $(1, \infty)$ to $(1 - \Delta, -1 + \Delta - \Delta \ln \Delta)$. Hence $-(m + h)$ increases from $-\infty$ to $\Delta \ln \Delta$, and $\lambda = 2\rho(-m-h) = \Delta^{-1}e^{-(m+h)/\Delta}$ increases from 0 to 1. For $\lambda < 1$, the avalanche size distribution has an exponential large-size tail

$$A_s(\lambda) \to \frac{1}{\lambda} \frac{1}{\sqrt{2\pi s^3}} \, e^{s(1-\lambda+\ln \lambda)} ,$$

which follows from (11.40) by Stirling's approximation. Approaching the bifurcation point where the infinite avalanche originates, $(m, h) = (m_*, -H_*)$ with $m_* = 1 - \Delta$ and $H_* = 1 - \Delta + \Delta \ln \Delta$, is equivalent to taking the $\lambda \to 1$ limit. In this situation, the large-size tail of the avalanche size distribution reduces to

$$A_s(\lambda) \simeq \frac{1}{\sqrt{2\pi s^3}} \, e^{-s(1-\lambda)^2/2}.$$

It is useful to rewrite this formula in terms of $h+H_*$, or $m-m_*$, rather than $1-\lambda$. Substituting $m = m_* + \epsilon$ and $h = -H_* + \delta$ into $m = 1 - e^{-(m+h)/\Delta}$ and expanding we obtain $\epsilon = \sqrt{2\Delta\delta}$ to leading order. Since $\lambda = 2\rho(-m - h)$ equals $\Delta^{-1}e^{-(m+h)/\Delta} = \Delta^{-1}(1 - m)$ for the exponential disorder distribution (11.34), we find

$$1 - \lambda = \frac{\epsilon}{\Delta} = \frac{m - m_*}{\Delta} = \sqrt{\frac{2\delta}{\Delta}} . \tag{11.41}$$

The large-size tail of the avalanche size distribution therefore becomes

$$A_s \simeq \frac{1}{\sqrt{2\pi s^3}} \, \epsilon^{-s(m-m_*)^2/2\Delta^2} = \frac{1}{\sqrt{2\pi s^3}} \, e^{-s(h+H_*)/\Delta} .$$

The average avalanche size $\langle s \rangle = \sum_{s \geq 1} s A_s \to (1 - \lambda)^{-1}$, i.e. it diverges as

$$\langle s \rangle \simeq \frac{\Delta}{m - m_*} = \sqrt{\frac{2\Delta}{h + H_*}} \ .$$

Infinite avalanche

Exactly at the bifurcation point $(m, h) = (m_*, -H_*)$ the avalanche size distribution has the algebraic tail

$$A_s = \frac{s^{s-1}}{s!} e^{-s} \simeq \frac{1}{\sqrt{2\pi s^3}} \ . \tag{11.42}$$

This distribution does not contain infinite avalanches and it is puzzling why the outcome (Fig. 11.4(b)) looks like an infinite avalanche that brings the system from $(m, h) = (m_*, -H_*)$ to $(m, h) = (-m_*, -H_*)$, with m_* found by solving $m_* = 1 - e^{-(m_* + H_*)/\Delta}$. To see how an infinite avalanche originates, note that, by decreasing the field through $-H_*$ so that the bifurcation point is passed, the magnetization will undergo many finite avalanches. After M such avalanches of sizes s_j, the magnetization decreases by

$$\delta m = \frac{2}{N} \sum_{j=1}^{M} s_j \ .$$

Each term in the sum is a random variable whose distribution A_s is given by (11.42). We estimate the largest term $s_{\max}(M)$ in this sum by the extremal criterion (see the highlight on page 17)

$$\sum_{s=s_{\max}(M)}^{\infty} A_s = \frac{1}{M} \ ,$$

which means that a value $\geq s_{\max}(M)$ occurs at most once in M trials. In conjunction with $A_s \sim s^{-3/2}$, this criterion gives $s_{\max}(M) \sim M^2$. Therefore $\sum s_j > s_{\max}(M) \sim M^2$. This lower bound for the largest term actually gives the correct asymptotic behavior of the entire sum. To show that this is the case, we write $\sum s_j = M \langle s \rangle_M$, where $\langle s \rangle_M$ is the mean value of the distribution A_s that is cut off at $s_{\max}(M)$. This gives[16]

$$\sum_{j=1}^{M} s_j \sim M \langle s \rangle_M = M \sum_{s=1}^{s_{\max}(M)} s A_s \sim M[s_{\max}(M)]^{1/2} \sim M^2 \ ,$$

from which $\delta m \sim M^2/N$. Let us estimate how M scales with N using the mathematical analogy between avalanches and components in the evolving random graph. First, we notice that an infinite avalanche arises with positive probability $A_\infty(\lambda) = 1 - \sum_{s \geq 1} A_s(\lambda)$,

---

[16] So the largest term therefore gives the correct scaling for the entire sum.

which scales as $2(\lambda - 1)$ for $\lambda \to 1$. On the other hand, using (11.41), we conclude that $\lambda - 1 \sim M^2/N$. The infinite avalanche will eventually be chosen when the number of avalanches M will be of the order of the inverse probability A_∞^{-1}. This gives $M \sim N/M^2$, or $M \sim N^{1/3}$. An infinite avalanche therefore emerges after $N^{1/3}$ finite avalanches.

11.5 Hysteresis in the random-field Ising chain

Consider the Hamiltonian of the RFIM in one dimension with (ferromagnetic) coupling constant $J = 1$,

$$\mathcal{H} = -\sum_i s_i s_{i+1} - \sum_i h_i s_i - h \sum_i s_i \,,$$

in which the external field decreases adiabatically. The absence of loops in the one-dimensional lattice greatly simplifies the analysis of hysteresis.[17] Initially $h = \infty$ and all spins point up. The spin at site i flips when its local field $s_{i-1} + s_{i+1} + h_i + h$ crosses zero. The state of the spin therefore depends not only on h_i and h, but also on the number of neighbors f that have already flipped. In one dimension, the number of such neighbors that are pointing down is $f = 0, 1$ or 2.

Let $\Pi_f(h)$ be the conditional probability that a spin with f spin-down neighbors is itself pointing down. Then

$$\Pi_f(h) = \int_{-\infty}^{-2(1-f)-h} \rho(x)\, dx \,. \tag{11.43}$$

When the external field decreases, the full history of spin-flip events can be complex. Fortunately, the Abelian and return-point memory properties of the ferromagnetic RFIM tremendously simplify the analysis. The return-point memory allows us to decrease the external field from ∞ to h in any way we like, e.g. suddenly, rather than doing it slowly and meticulously recording avalanche histories. The Abelian property allows us to flip spins that are anti-aligned with their local fields in any order that we prefer.

Thus let's start with all the spins pointing up ($h = \infty$) and suddenly decrease the external field to h. We now ask: what is the probability $p(h)$ that an arbitrary *seed* spin flips? By the Abelian property, we compute this probability under the assumption that the seed spin is the last one to flip. The probability that the seed spin flips depends on the state of its neighbors. Denote by $\Pi(h)$ the probability that the right neighbor of the seed spin (that we call the R spin) flips. Then

$$p(h) = \left[\Pi(h)\right]^2 \Pi_2(h) + 2\Pi(h)\left[1 - \Pi(h)\right]\Pi_1(h) + \left[1 - \Pi(h)\right]^2 \Pi_0(h) \,. \tag{11.44}$$

For example, the factor $\Pi^2(h)$ in the first term accounts for both the right and left neighbors of the seed spin pointing down. Since the seed spin has been constrained to point up, the

---

[17] A similar simplification arose in disordered spin chains in Section 10.1.
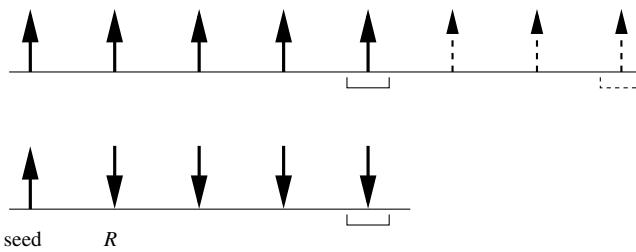
Fig. 11.6. Avalanche that causes the seed spin to flip. After the field is decreased, two flippable spins are created (bracketed), but all dashed spins can be ignored. The closest flippable spin creates an avalanche of size 4 that propagates to the left and reaches R.

semi-infinite chains to the left and right of the seed spin are uncorrelated, so that $\Pi^2(h)$ appears. We then multiply $\Pi^2(h)$ by the corresponding conditional probability $\Pi_2(h)$ to give the first term. The other terms have similar explanations.

Determining $\Pi(h)$ is conceptually simpler than finding $p(h)$ because the problem is one-sided. Indeed, the R spin can flip only due to an avalanche that propagates from right to left or originates at the R spin itself. When the external field has suddenly decreased from ∞ to h, a fraction $\Pi_0(h)$ of initially flippable spins is created (Fig. 11.6). We may ignore all spins to the right of the closest flippable spin (dashed in the figure), because they do not influence the fate of the R spin. When the closest flippable spin flips, it initiates an avalanche that propagates to the left and reaches R. By the Abelian property, we can perform this avalanche sequence in right-to-left order.

By tracking only this finite number of relevant spin flips, their probabilities are readily computable. For example, for the configuration of Fig. 11.6, the probability for this avalanche of size 4 is $\Pi_0(h)[\Pi_1(h) - \Pi_0(h)]^3$. The factor $\Pi_0(h)$ gives the probability for the closest flippable spin that starts the avalanche and $[\Pi_1(h) - \Pi_0(h)]^3$ gives the probability that each of the next three spins will flip when none of them was initially flippable. Summing over all avalanche sizes $l \geq 1$ gives

$$\Pi(h) = \sum_{l=1}^{\infty} \Pi_0(h)\big[\Pi_1(h) - \Pi_0(h)\big]^{l-1} = \frac{\Pi_0(h)}{1 - \big[\Pi_1(h) - \Pi_0(h)\big]} \ . \tag{11.45}$$

As is often the case in combinatorial arguments, the result of Eq. (11.45) is simple, yet its derivation is far from obvious. Therefore we give another short, but abstract, probabilistic derivation. Owing to translational invariance, the right neighbor of the R spin also flips with probability $\Pi(h)$. Then there are two contributions to the probability $\Pi(h)$ that the R spin will flip: $\Pi(h)\,\Pi_1(h)$ when its right neighbor flips, and $[1 - \Pi(h)]\,\Pi_0(h)$ when it does not. Thus

$$\Pi(h) = \Pi(h)\,\Pi_1(h) + [1 - \Pi(h)]\,\Pi_0(h)\,, \tag{11.46}$$

which is equivalent to (11.45).

Equations (11.44)–(11.45) give the general solution for an arbitrary disorder distribution. We compute $\Pi_1(h) = \int_{-\infty}^{-h} \rho(x)\,dx$ directly; then the two other conditional probabilities

are found from $\Pi_0(h) = \Pi_1(h+2)$ and $\Pi_2(h) = \Pi_1(h-2)$. Substituting these $\Pi_f(h)$ into Eqs (11.44)–(11.45), we find $p(h)$. Finally, the magnetization is given by $m(h) = 1 - 2p(h)$. Let's specialize the solution to the case of the uniform disorder distribution (11.32). We find

$$\Pi_1(h) = \begin{cases} 1, & h < -\Delta, \\ (\Delta - h)/(2\Delta), & |h| \leq \Delta, \\ 0, & h > \Delta. \end{cases}$$

When the disorder is weak ($\Delta < 1$), the hysteresis loop is rectangular, as in the case of the mean-field RFIM with a uniform disorder distribution (Fig. 11.4(a)). The only difference is the magnitude of the coercive field, $H_* = 2 - \Delta$ rather than $H_*^{MF} = 1 - \Delta$. In contrast to the mean-field case, the static hysteresis loop does not disappear at the critical strength $\Delta = 1$. Instead, its shape changes when Δ exceeds $\Delta_1 = 1$ and it undergoes another bifurcation when the disorder strength exceeds $\Delta_2 = 2$. For $\Delta > 2$, the upper branch of the hysteresis loop is (see Fig. 11.7 and problem 11.8)

$$m(h) = \begin{cases} -1, & h < -\Delta, \\ \dfrac{\Delta(6\Delta - 4 - \Delta^2) + \Delta(5\Delta - 4)h + (\Delta - 2)h^2 - h^3}{4\Delta(\Delta - 1)^2}, & -\Delta < h < 2 - \Delta, \\ \dfrac{2 + (\Delta + 1)h}{\Delta(\Delta - 1)}, & 2 - \Delta < h < \Delta - 2, \\ 1, & h > \Delta - 2. \end{cases}$$

(11.47)

For $1 < \Delta \leq 2$, no part of the magnetization curve has a linear dependence on the external field, and $m = m(h)$ is a cubic polynomial when $-\Delta < h < -2 + \Delta$. The primary difference with the mean-field RFIM is that the hysteresis loop does not disappear when the disorder becomes large. This result is generic, while more subtle features of the hysteresis loop do depend on the disorder distribution.
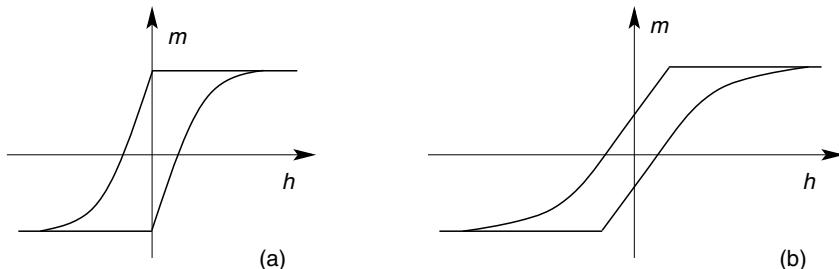
Fig. 11.7. The hysteresis loop for the uniform disorder distribution. (a) $\Delta = 2$. The non-trivial part of the upper branch is $m = 1 + \frac{3}{2} h - \frac{1}{8} h^3$ for $-2 < h < 0$. (b) $\Delta = 3$. The magnetization is linear, $m = \frac{1}{3} + \frac{2}{3} h$ when $-1 < h < 1$; when $-3 < h < -1$, the magnetization is $m = \frac{5}{16} + \frac{11}{16} h + \frac{1}{48} h^2 (1 - h)$.

Hysteresis on Cayley trees

We finish with a few comments regarding hysteresis in the RFIM on a Cayley tree. There is some formal similarity between the RFIM on the Cayley tree and in one dimension because both systems do not have loops. The absence of loops allows one to solve, for example, the random sequential adsorption of dimers on a Cayley tree by using the same technique as that in one dimension (see Chapter 7). However, even the equilibrium properties of the ferromagnetic RFIM on the Cayley tree, such as the average free energy or the magnetization, are unknown, even at zero temperature and in zero external uniform field.[18] In spite of this caveat, the hysteresis loop in the RFIM on the Cayley tree can be determined analytically for an *arbitrary* continuous distribution of random fields.

The line of attack involves the straightforward generalization of Eq. (11.44) that applies for the one-dimensional case. For a seed spin, the probability $p(h)$ that it will flip is related to the probability $\Pi(h)$ that one of its neighbors will flip by

$$p = \sum_{f=0}^{z} \binom{z}{f} \Pi^f \, (1 - \Pi)^{z-f} \, \Pi_a \,, \qquad (11.48)$$

where $\Pi_f(h) = \int_{-\infty}^{-(z-2f)-h} \rho(x)\,dx$ are the conditional probabilities, and $z \geq 2$ is the coordination number of the Cayley tree. Similarly, the generalization of Eq. (11.46) is

$$\Pi = \sum_{f=0}^{z-1} \binom{z-1}{f} \Pi^f \, (1 - \Pi)^{z-1-f} \, \Pi_f \,. \qquad (11.49)$$

In computing the probability $\Pi(h)$ that the neighboring spin j of the seed spin flips, we have to specify the state of the remaining $z - 1$ neighbors of spin j. The binomial factor $\binom{z-1}{f}$ accounts for the number of configurations with f down spins and $z - 1 - f$ up spins, $\Pi^f \, (1 - \Pi)^{z-1-f}$ gives the probability of this configuration, and Π_f is the conditional probability for the spin j to flip.

Equations (11.48)–(11.49) give the general solution for the hysteresis loop in the RFIM on the Cayley tree that is valid for an arbitrary disorder distribution and arbitrary coordination number. The full analysis of the magnetization loop $m(h)$ that is obtained by solving (11.48)–(11.49) is cumbersome, even for the uniform disorder distribution. The outcome, however, is simple and surprising. For $z = 3$ the magnetization loops are qualitatively similar to those found for $z = 2$, i.e. in one dimension. In contrast, for $z \geq 4$, the magnetization loops are qualitatively similar to those found in the mean-field model. This result is not surprising, since the Cayley tree is effectively infinite-dimensional. What is odd, however, is that the hysteresis characteristics of the Cayley tree for the case $z = 3$ are quasi-one-dimensional.

---

[18] The ground-state properties of the RFIM on Cayley trees have been computed only for a bimodal distribution of random fields, $\rho(x) = \frac{1}{2}\,\delta(x-1) + \frac{1}{2}\,\delta(x+1)$.

11.6 Notes

Hysteresis has been vital in tape recording and other magnetic-based storage technologies. For example, the basic property that the magnetization of a tape lags behind the field from the tape head has been used in storing music. Nevertheless, hysteresis is often viewed as a nuisance that should be eliminated to obtain "correct" equilibrium data. On the theoretical side, hysteresis involves many challenges. A proper description of hysteresis requires non-equilibrium techniques, taking disorder into account (the common term "dirty materials" probably better reflects their essence than the scientific name "disordered systems"), and possibly accounting for the effects of long-ranged magnetic and even elastic interactions.

Our presentation of hysteresis focused on two special models. The first, even if phrased in magnetic language (a homogeneous ferromagnet in an oscillating magnetic field), is quite general and it can be adapted to other dynamical systems that are subjected to an oscillating external force. For more information about this type of model see [199, 200]. Scaling laws in dynamical hysteresis are discussed in [201, 202], which also use asymptotic expansions. The book by Bender and Orszag [138] is an excellent reference on these types of asymptotic techniques. For an exposition of the Poincaré map and elementary topological methods used in analysis of ordinary differential equations consult Strogatz [203].

The argument for the stability of the ferromagnetic state in the presence of random magnetic fields is due to Imry and Ma [204]. Feynman [205], in chapter 37 of volume II, gives a readable account of the jerky magnetization curve as a function of magnetic field. The RFIM that accounts for this phenomenon was suggested and analyzed by Sethna and co-workers; see the review in [206] and also the book [207]. This model admits an analytical treatment in one dimension, and also on Cayley trees; our presentation was based on [208–210]. For reviews of hysteresis in magnetic materials, see [206, 211].

11.7 Problems

11.1 Show that for the mean-field ferromagnet, the magnitude of the coercive field is given by (11.5).
Hint: To determine H_* and $m_* = \sqrt{1-T}$ use the validity of (11.4) at the point $(-m_*, H_*)$ in the m–H plane, together with the fact that at $(-m_*, H_*)$ the curve $F(m) = \tanh[(m+H)/T]$ has a slope equal to 1 when this curve touches the diagonal.

11.2 Show that the second term in the expansion (11.10) is given by

$$m_2 = -\left[\frac{dm_1}{d\tau} + m_1^2\,\frac{m_0(1-m_0^2)}{T^2}\right]\left[1 - \frac{1-m_0^2}{T}\right]^{-1},$$

where m_0 and m_1 are found from (11.13).

11.3 Show that the second term in the expansion (11.15) satisfies

$$\frac{dm_2}{d\tau} = -m_1 + \frac{m_1}{T}\left[1 - \tanh^2\left(\frac{H\sin\tau}{T}\right)\right]$$

and that the above equation admits a periodic solution if

$$\int_0^{2\pi} m_1(\tau)\,d\tau = \int_0^{2\pi} \frac{m_1(\tau)}{T}\left[1 - \tanh^2\left(\frac{H\sin\tau}{T}\right)\right]d\tau.$$

11.4 The goal of this problem is to show that the constant C that appears in (11.25) is well-defined and positive, i.e. that the integral $I_0 = \int_{-\infty}^{\infty} M_0(\theta)\,d\theta$ exists and is negative.

(a) Establish the divergence of $M_0(\theta)$ for large $|\theta|$. Namely, using (11.23) deduce the asymptotic

$$M_0 \to (3\theta)^{1/3} - (3\theta)^{-4/3} \qquad \text{as} \qquad |\theta| \to \infty.$$

(b) Establish the validity of the relation $\int_{-\infty}^{\infty}(3\theta)^{1/3}\,d\theta = 0$. (The integrals $\int_{-\infty}^{\infty}(\ldots)\,d\theta$ are defined as $\lim_{\Theta\to\infty}\int_{-\Theta}^{\Theta}(\ldots)\,d\theta$.)

(c) Using the above relation rewrite I_0 as the integral $I_0 = \int_{-\infty}^{\infty} w(\theta)\,d\theta$ with $w(\theta) = M_0(\theta) - (3\theta)^{1/3}$. Show that the asymptotic behavior of the function $w(\theta)$ implies that it is integrable, and hence the integral I_0 exists.

(d) To establish that the integral I_0 is negative it suffices to show that $w(\theta)$ is negative. This fact can be derived as follows:

   (i) Recall that $w(\theta)$ is negative for $\theta \to -\infty$.

   (ii) Using this previous assertion show that, if $w(\theta)$ is somewhere positive, there would be a first point ϑ where $w(\vartheta) = 0$. Argue that the derivative $w'(\vartheta)$ at this point should be non-negative, since the graph of $w(\theta)$ crosses $w = 0$ from below at point ϑ.

   (iii) Show by direct computation that

$$\frac{dw}{d\theta} = \frac{dM_0}{d\theta} - (3\theta)^{-2/3} = \theta - \frac{1}{3}M_0^3 - (3\theta)^{-2/3}.$$

Using this result, in conjunction with the presumed relation $M_0(\vartheta) = (3\vartheta)^{1/3}$, deduce that the derivative at point ϑ is $w'(\vartheta) = -(3\vartheta)^{-2/3}$, i.e. it is negative. This contradiction proves that w is indeed negative.

11.5 This problem is concerned with hysteresis in the ferromagnetic Ising chain in the limit of small field: $H \ll \min(J, T)$.

(a) Show that in the low-frequency limit, $\Omega \ll 1 - \gamma$, Eq. (11.30) simplifies to

$$\chi \equiv \frac{M}{H} = \frac{1}{T}\frac{1 - \eta^2}{1 + \eta^2}\frac{1}{1 - \gamma} = \frac{1}{T}\frac{1 - \eta}{1 + \eta} = \frac{1}{T}\exp\left(\frac{2J}{T}\right).$$

(b) Verify that the above expression coincides with the standard expression for the static susceptibility $\chi = M/H$ of the ferromagnetic Ising chain.

11.6 Consider the mean-field RFIM with coupling constant $1/N$ between all pairs of spins at zero temperature.

(a) Show that for the uniform disorder distribution (11.32), the magnetization is given by (11.33) when the disorder is sufficiently strong.

(b) Establish the asymptotic behavior (11.37) of magnetization for the exponential disorder distribution.

(c) Consider the Gaussian distribution

$$\rho(h_i) = (2\pi \Delta^2)^{-1/2} \exp\left(-\frac{|h_i|^2}{2\Delta^2}\right)$$

of the random fields. Verify that the critical disorder strength is $\Delta_c = \sqrt{2/\pi}$. Show that when the hysteresis loop is absent, $\Delta \geq \Delta_c$, the magnetization exhibits the following asymptotic behavior:

$$m = \begin{cases} \left(\frac{\Delta}{\Delta_c} - 1\right)^{-1} h + \mathcal{O}(h^2), & \Delta > \Delta_c, \\ (3\pi/4)^{1/3} h^{1/3} + \mathcal{O}(h^{2/3}), & \Delta = \Delta_c. \end{cases}$$

(d) Establish the asymptotic behaviors (11.38) and (11.39).

11.7 Continue the line of reasoning discussed on page 364 to determine the avalanche size distribution $A_s(\lambda)$ quoted in Eq. (11.40) for the mean-field RFIM. What is needed is the probability that exactly s spins can flip in a single avalanche. For this event to occur, the values of the random fields for these s spins have to lie in the range $[x, x + 2s/N]$, and the distance between successive random-field values cannot be larger than $2/N$.

11.8 Derive the upper branch of the hysteresis loop for the uniform distribution of the disorder with $\Delta > 2$, Eq. (11.47).

Population dynamics

The dynamics of interacting populations are usually modeled by coupled nonlinear differential equations. In this chapter, we will outline some of the basic phenomenology of a few classic models of population dynamics. As we shall see, stochasticity can, surprisingly, play an essential role even in the simplest situations. Consequently, we also present statistical physics approaches that yield insights about the phenomena that lie outside a continuum description of the dynamics, such as large fluctuations and the existence of different asymptotic states for the continuum and stochastic dynamics.

12.1 Continuum formulation

Consider a system that contains N different species. We assume that all species are perfectly mixed throughout the evolution so that the basic variables are their global densities ρ_i. The evolution of these densities in the continuum limit is described by the set of equations
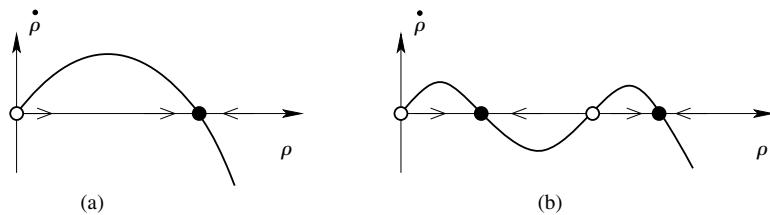
$$\dot{\rho}_i = F_i(\boldsymbol{\rho}) \,.$$

Here the overdot denotes the time derivative and $\boldsymbol{\rho} = (\rho_1, \rho_2, \rho_3, \ldots)$ denotes the set of species densities. The functions F_i encode the reactions and are typically nonlinear functions of the ρ_i. For more than two species, the resulting dynamical behavior can be arbitrarily complex – including strange attractors, topologically complex orbits, and chaos. Our focus is on simple systems for which these complications do not arise, but for which stochasticity effects can be important.

Logistic growth

We begin with single-species *logistic growth* that describes the evolution of a population of self-replicating and mutually competing organisms. Individuals reproduce asexually (rate r) and die due to competition with other members of the population (rate k). The density ρ evolves according to $\dot{\rho} = r\rho - k\rho^2$. This equation can be written in dimensionless form by rescaling the time $t \to rt$ and the density by $\rho \to k\rho/r$ to give the *logistic equation*

$$\dot{\rho} = \rho(1 - \rho) \,, \tag{12.1}$$

Fig. 12.1. Phase portraits of: (a) logistic growth, and (b) a more general reaction $\dot\rho = F(\rho)$. The open and solid circles denote unstable and stable fixed points, respectively. The arrows indicate the direction of the flow of ρ.

whose solution is

$$\rho(t) = \frac{\rho(0)\, e^t}{1 + \rho(0)(e^t - 1)}\,. \tag{12.2}$$

Thus any initial density evolves exponentially quickly to the steady-state value $\rho = 1$.

More generally, consider reactions that are governed by $\dot\rho = F(\rho)$, where $F(\rho)$ has multiple extrema and zeros. We can understand the dynamical behavior with minimal calculation by studying the *phase portrait* of the system, namely a plot of $\dot\rho$ as a function of ρ (Fig. 12.1). The points where $F = 0$ are *fixed points* of the dynamics. In the long-time limit, a single-species system must eventually converge to one of these fixed points. There are two types of fixed points – *stable* and *unstable*. If the system starts at an unstable fixed point, it will remain there forever. If the system starts infinitesimally close to an unstable fixed point, it will eventually reach a stable fixed point. The range of initial states that ends at a given fixed point is the *domain of attraction* of this fixed point. If we are unable to solve the equation $\dot\rho = F(\rho)$ explicitly (and usually we cannot), we can still understand its dynamical behavior from the phase portrait. We merely find all the fixed points ρ^*, where $F(\rho^*) = 0$, and determine their stability by the sign of F' at each fixed point. Then we can infer the long-time behavior from the direction of the "flow," as illustrated in Fig. 12.1. In this way, the dynamical behavior of any one-dimensional dynamical system can be viewed as solved.

Example 12.1. *Approach to the fixed point*. Consider the equation $\dot\rho = F(\rho)$ close to a stable fixed point where $F(\rho^*) = 0$. Expanding $F(\rho) = -\lambda(\rho - \rho^*) + \cdots$ gives the equation of motion, $\dot\rho = -\lambda(\rho - \rho^*)$, infinitesimally close to the fixed point. The solution is $\rho - \rho^* \sim e^{-\lambda t}$. If, however, $F'(\rho^*)$ also vanishes, then the expansion of F about a stable fixed point is now $F(\rho) = -\lambda(\rho - \rho^*)^2 + \cdots$. The local equation of motion becomes $\dot\rho = -\lambda(\rho - \rho^*)^2$, with solution $\rho - \rho^* \simeq (\lambda t)^{-1}$. To summarize, when $F'(\rho^*)$ is non-zero, the fixed point is approached exponentially quickly. If the first term in the expansion of F near a fixed point is faster than linear, then the fixed point is approached as a power law in time.

In the framework of population dynamics, there are many ways to generalize logistic-type models. As examples, one can augment the equation $\dot\rho = F(\rho)$ by incorporating delayed

response, a population age structure, or having discrete and non-overlapping generations rather than continuous time. These generalizations effectively make the dynamical system higher- (or infinite-) dimensional and the evolution of such systems can be quite rich.

Two-species reactions

For two distinct reactive species A and B, the range of reaction possibilities expands considerably. We discuss three generic examples:

- competition, where A and B annihilate each other;
- symbiosis, where A and B reinforce each other;
- predator–prey, where A feeds on B.

For the intra-species reactions, it is natural to consider logistic growth. However, for the predator–prey reaction, we will study a simpler intra-species interaction to highlight a unique feature of this system.

Competition and symbiosis

In two-species competition, each species undergoes logistic growth that is augmented by a competitive interaction. The evolution of the densities, now also denoted by A and B, is described by

$$
\begin{aligned}
\dot{A} &= A(1 - A - \epsilon B) \equiv F_A(A, B)\,, \\
\dot{B} &= B(1 - B - \epsilon A) \equiv F_B(A, B)\,,
\end{aligned}
\tag{12.3}
$$

where the magnitude of the competition is quantified by ϵ. For simplicity, we assume that each species has the same dynamics in the absence of competition and that the effect of the competition is also symmetric (non-symmetric intra-species interactions are the subjects of problems 12.1 and 12.2). To understand the long-time behavior of this system, let's first find the fixed points where $\dot{A} = \dot{B} = 0$. These are located at

$$
(A^*, B^*) = (0, 0), \qquad (1, 0), \qquad (0, 1), \qquad \text{and} \quad (1/(1 + \epsilon), 1/(1 + \epsilon))\,.
$$

The last fixed point is the interesting one, as it describes the effect of competition on coexistence. To learn if this fixed point actually describes the effect of the competition, we need to determine its stability. For the two-dimensional dynamical system (12.3), the analogs of the derivative at the fixed point are the eigenvalues and eigenvectors of the linearized rate equations at the fixed point:

$$
\begin{pmatrix}
\dfrac{\partial F_A}{\partial A} & \dfrac{\partial F_A}{\partial B} \\[2ex]
\dfrac{\partial F_B}{\partial A} & \dfrac{\partial F_B}{\partial B}
\end{pmatrix}.
\tag{12.4}
$$

The eigenvalues of the matrix (12.4) at each of the four fixed points are:

$$
\begin{array}{lll}
\text{at } (0,0) & \lambda_1 = 1, & \lambda_2 = 1, \\
\text{at } (1,0) & \lambda_1 = -1, & \lambda_2 = 1 - \epsilon, \\
\text{at } (0,1) & \lambda_1 = 1 - \epsilon, & \lambda_2 = -1, \\
\text{at } (1/(1 + \epsilon), 1/(1 + \epsilon)) & \lambda_1 = 1, & \lambda_2 = (\epsilon - 1)/(\epsilon + 1).
\end{array}
$$

For the fixed point $(0,0)$ both eigenvalues are positive; thus the flow diverges from this fixed point and $(0,0)$ is an unstable fixed point. For the other three fixed points, there can be two negative eigenvalues, which means that the fixed point is stable, or one positive and one negative eigenvalue. This latter case means that the flow converges to the fixed point in one direction and diverges in an orthogonal direction. Such fixed points are known as *saddle* points.

It is generally easier and physically more clear to determine the global flow by focusing on the *nullclines*, namely, the loci for which $\dot{A} = 0$ and $\dot{B} = 0$. Setting $\dot{A} = 0$ and $\dot{B} = 0$, the nullclines are given by

$$
A = 1 - \epsilon B, \qquad B = 1 - \epsilon A,
$$

which are plotted in Fig. 12.2 for both $\epsilon < 1$ and $\epsilon > 1$, corresponding to weak and strong competition, respectively.[1] The intersection of the nullclines defines a fixed point. Since we know the flow direction of a single density on either side of a single nullcline, we can visually infer the global flow in the A–B plane from the combined information of the two nullclines. We thereby find that for weak competition, the fixed point where both densities coexist is stable, albeit at a reduced level compared to the case of no competition. For strong competition, however, any density difference between the two species is magnified and only the stronger survives in the long-time limit.
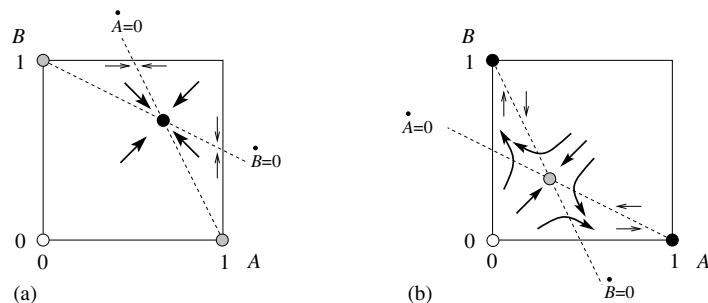
Fig. 12.2. The A–B density plane for two-species competition (Eq. (12.3)), showing the nullclines $\dot{A} = 0$ and $\dot{B} = 0$ (dashed) for (a) $\epsilon < 1$ and (b) $\epsilon > 1$. The small arrows near a given nullcline indicate the direction of flow of the corresponding density. Also shown are the fixed points, with open, shaded, and solid circles for unstable, saddle, and stable fixed points, respectively. The heavy arrows indicate the flow of A and B.

---

[1] There are two more nullclines $A = 0$ and $B = 0$, but these play no role when both densities are non-zero.

By allowing the parameter ϵ in Eq. (12.3) to be negative, we can also treat symbiosis, where the presence of one species reinforces the other. Defining $\delta = -\epsilon$, the coexistence fixed point is now located at

$$(A^*, B^*) = (1/(1 - \delta), 1/(1 - \delta)).$$

From the linearized equations about each fixed point, there are two possible behaviors. For $\delta < 1$, the symbiosis is sufficiently mild that the steady-state densities are merely increased compared to the case of no symbiosis. However, for $\delta > 1$, the coexistence fixed point is unstable and the densities grow without bound for most initial conditions.

Predator–prey interactions

A particularly appealing population dynamics model was proposed by Volterra in 1926 to explain the observed oscillations of fish catches in the Adriatic sea. The basic idea was that the population consisted of a prey species, that would reproduce freely in the absence of predators, and a predator that eats the prey to survive. Let's write the simplest dynamical equations that account for these two processes:
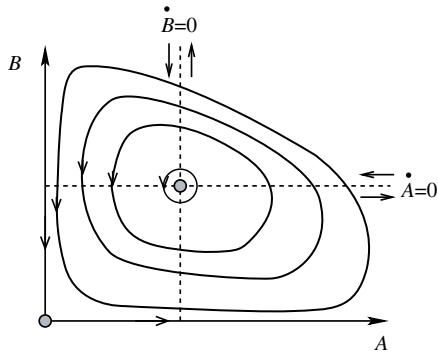
$$\begin{aligned} \dot{A} &= A - AB, \\ \dot{B} &= -B + AB. \end{aligned} \tag{12.5}$$

Here A and B correspond to the prey and predator densities, respectively. The system is known as the *Lotka–Volterra* model. To eliminate notational clutter, we set the four potentially distinct rate constants on the right-hand sides to one. Without loss of generality, it is possible to eliminate three of these constants by suitably rescaling the time and the two densities, so that the general model has only a single rate constant. For simplicity, we also dispense with a quadratic self-regulation term for each species (this generalization is the subject of problem 12.2). From Eq. (12.5), the prey density A would grow exponentially in time if the world were free of predators B. However, the prey are eaten at a rate that is proportional to encounters between prey and predator. Similarly, predators would starve and disappear exponentially in time in the absence of prey, but are nourished by eating them.

The two fixed points of this system are $(A^*, B^*) = (0, 0)$ and $(1, 1)$. Performing a linear stability analysis about each fixed point, we find that the fixed point at the origin is a saddle – stable in the vertical direction and unstable in the horizontal direction – while the fixed point at $(1, 1)$ is a *neutral center*. A center means that both eigenvalues are complex so that, infinitesimally near a center, the flow is rotational. For an unstable center, the flow diverges because the eigenvalues have a positive real part, and vice versa for a stable center. For a neutral center, the real parts of the eigenvalues are zero and the flow is purely circular. It is again simple to patch together the global behavior by studying the nullclines. Setting $\dot{A} = 0$ and $\dot{B} = 0$, the nullclines are given by

$$A(1 - B) = 0, \qquad B(A - 1) = 0,$$

from which we infer that the global flow is rotational about the coexistence fixed point (Fig. 12.3). In fact, all orbits are closed, a feature that stems from a hidden conservation law

Fig. 12.3. Lotka–Volterra model in the A–B plane. Dashed lines indicate the nullclines, and the closed curves represent periodic orbits for a given value of the conserved quantity $\mathcal{C}$ in Eq. (12.6).

in the Lotka–Volterra model. Any modifications of the equations of motion (12.5) to make the model more realistic, such as including self-regulation for each species, invalidates the conservation law and changes the global nature of the flow.

To find the conservation law, we divide the first of (12.5) by A and the second by B, to give equations for $\ln A$ and $\ln B$:
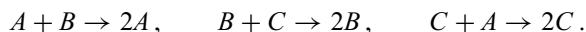
$$\frac{d \ln A}{dt} = 1 - B, \qquad \frac{d \ln B}{dt} = A - 1.$$

Adding these two equations, the right-hand side equals $A - B$, which is the same, from Eq. (12.5), as $\dot{A} + \dot{B}$. Thus the conservation law is

$$\ln(A\,B) - (A + B) \equiv \mathcal{C} = \text{const.} \tag{12.6}$$
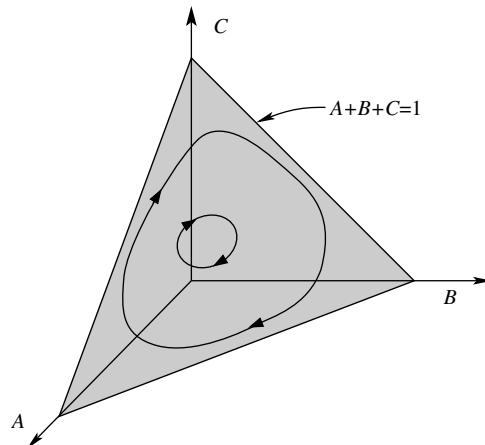
If the initial state is not $(0, 0)$ or $(1, 1)$, then the subsequent motion is along a closed orbit that is characterized by a given value of $\mathcal{C}$. Each of these orbits is marginally stable, which means that a sudden infinitesimal perturbation moves the system from one orbit to another nearby orbit, where it would remain forever in the absence of additional perturbations.

Example 12.2. *Cyclic competition*. A beautiful example of predator–prey interaction with more than two species is the model dubbed "the struggle for existence" by G. F. Gause (Fig. 12.4). Here the three species struggle for supremacy via the cyclic set of interactions

$$A + B \rightarrow 2A, \qquad B + C \rightarrow 2B, \qquad C + A \rightarrow 2C.$$

The densities of these three species, A, B, and C, evolve according to

$$\begin{aligned}
\dot{A} &= A(B - C), \\
\dot{B} &= B(C - A), \\
\dot{C} &= C(A - B).
\end{aligned} \tag{12.7}$$

Fig. 12.4. Dynamics of "the struggle for existence" model, Eq. (12.7), in the *ABC* plane. Shown are two periodic orbits that represent the intersection of the plane $A + B + C = 1$ with the hyperboloid $ABC =$ constant.

This symmetric set of equations admits two conservation laws. First, by adding Eqs (12.7), we see that the total density is conserved, $A + B + C = 1$. Next, we write the equations for $\dot{A}/A$, $\dot{B}/B$, and $\dot{C}/C$, and add them to find that the product of the densities ABC is also conserved. Thus for a given value of the constant $\mathcal{C} \equiv ABC$, the evolution is confined to a closed orbit in the plane defined by $A + B + C = 1$.

Example 12.3. *Three-species predation.* As an example of the complexities that can arise in a three-species system, consider a predator–prey model in which each species undergoes logistic growth and also preys on the other two species cyclically. The evolution of this system is described by

$$
\begin{aligned}
\dot{A} &= A(1 - A - \alpha B - \beta C)\,, \\
\dot{B} &= B(1 - B - \alpha C - \beta A)\,, \\
\dot{C} &= C(1 - C - \alpha A - \beta B)\,,
\end{aligned}
\tag{12.8}
$$

where α and β quantify the intensity of the predation between unlike species. There are three distinct asymptotic outcomes of this dynamics as a function of the competition parameters α and β.

- *Weak predation*, $\alpha + \beta < 2$. The fixed point where the densities of all three species exist is stable.
- *Strong predation*, $\alpha > 1$ and $\beta > 1$. The fixed points where only a single species exists are stable. Which of the fixed points turns out to be the attractor of the dynamics – either $(1,0,0)$, $(0,1,0)$, or $(0,0,1)$ – depends on the initial condition.
- *Asymmetric predation*. For all other α and β values, there is no stable fixed point and the evolution is dynamically complex.

The latter case is particularly intriguing; it can be shown that the density of each species oscillates in a non-periodic way, with the period proportional to the total elapsed time (problem 12.4).

Epidemics

The spread of epidemics can be formulated as a set of reactions in which populations of healthy, susceptible, and infected individuals interact and propagate a disease. Two classic models for epidemic spread are the *SIS* (susceptible–infected–susceptible) and the *SIR* (susceptible–infected–recovered) models. The former represents a caricature for the dynamics of the common cold, in which an individual can experience endless cycles of infection and recovery. The latter describes a disease that an individual gets only once in a lifetime, after which either full immunity is conferred or the individual dies – both outcomes may be lumped into the category of "remove."

For the *SIS* model, we adopt the notation that the densities of susceptible and infected individuals are S and I, respectively. When a susceptible and an infected individual meet, the susceptible becomes infected at a rate k that quantifies the virulence of the infection. Conversely, an infected individual spontaneously recovers with rate r. These densities evolve according to

$$\begin{aligned} \dot{S} &= rI - kSI\,, \\ \dot{I} &= -rI + kSI\,. \end{aligned} \tag{12.9}$$

The *SIS* model is nothing more than logistic growth in disguise. From Eqs (12.9), the total density $S + I$ (that we may set to 1) is conserved. Thus using $S = 1 - I$, the evolution equation for I is

$$\dot{I} = (k - r)I - kI^2\,, \tag{12.10}$$

which is just the logistic equation (12.1), but with specified rates. If the infection rate k exceeds the recovery rate r, then the infected density approaches the stationary value $1 - r/k$, with the same exponential time dependence as in Eq. (12.2). Correspondingly, the density of susceptible individuals approaches the stationary value r/k. Conversely, if $k < r$, then $I \to 0$ while $S \to 1$. In both cases, the infected density S is "slaved" to I; that is, once I is determined, then S is also determined.

The *SIR* dynamics is more interesting. While the infection step is the same as in the *SIS* model, recovered individuals are effectively removed from the population and their fraction R can never decrease. The densities of susceptible, infected, and recovered individuals evolve according to

$$\begin{aligned} \dot{S} &= -kSI\,, \\ \dot{I} &= -rI + kSI\,, \\ \dot{R} &= rI\,. \end{aligned} \tag{12.11}$$

The basic question of the *SIR* model is this: Can a small fraction of infected individuals in an otherwise susceptible population trigger a macroscopic epidemic? To determine the criterion for an epidemic, notice, from the first of Eqs (12.11), that the density of susceptible individuals can only decrease with time. Then for $S < r/k \equiv \bar{r}$, the second equation, $\dot{I} = I(kS - r)$, shows that the density of infected individuals decreases with time – there is no outbreak. Since S decreases with time, if there is no outbreak initially, there can never be an outbreak.

Conversely, for $S > \bar{r}, \dot{I}$ will initially be positive and an epidemic is initiated. To determine the severity of the epidemic, we divide the second of Eqs (12.11) by the first to give
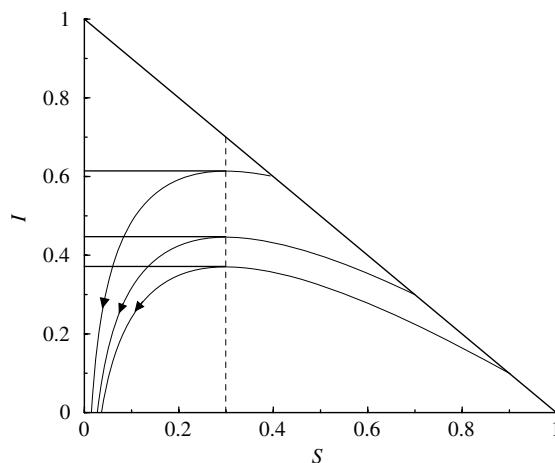
$$\frac{dI}{dS} = -1 + \frac{\bar{r}}{S} . \tag{12.12}$$

Integrating this equation, the values of S and I on a given trajectory in the S–I plane are related by

$$I + S - \bar{r} \ln S = I(0) + S(0) - \bar{r} \ln S(0) . \tag{12.13}$$

From Eq. (12.12), it is only for $S(0) > \bar{r}$ that an epidemic can arise in which the density of infected individuals initially increases (Fig. 12.5). The maximum density of infected individuals at the peak of the epidemic, I_{max}, occurs when $S(t) = \bar{r}$. Using this fact in Eq. (12.13), we obtain

$$I_{max} = 1 - \bar{r} + \bar{r} \ln \left[\frac{\bar{r}}{S(0)} \right] . \tag{12.14}$$

Fig. 12.5. Evolution of the *SIR* model in the *S–I* plane. The trajectories shown correspond to starting at $S(0) = 0.9$, 0.7, and 0.4 (and $I(0) = 1 - S(0)$), respectively, for the case $\bar{r} = 0.3$ (vertical dashed line). The horizontal lines give the maximum infected fraction I_{max} for the three cases shown. The intersection of each trajectory with the horizontal axis gives $S(\infty)$.

Another important characterization is the fraction of individuals I_{tot} that have contracted the disease at any time during the epidemic. Using again Eq. (12.13), but now in the infinite-time limit where $I \to 0$ and $S \to S(\infty)$, we obtain the transcendental equation for $S(\infty)$,

$$S(\infty) = S(0)e^{-[1-S(\infty)]/\bar{r}} \, , \tag{12.15}$$

from which $I_{\text{tot}} = 1 - S(\infty)$. For example, from Eq. (12.15), if the population initially consists of 90% susceptible individuals and 10% infected individuals and $\bar{r} = 0.3$ as in Fig. 12.5, then the maximum fraction of infected individuals at the peak of the epidemic is approximately 37% and the fraction of individuals that were infected before the epidemic has burned itself out is approximately 96% – almost everyone will have gotten sick!

Example 12.4. *Rumor spreading*. A simple variant of the *SIR* model describes rumor spreading through a population that consists of spreaders (S), stiflers (T), and those who are unaware (U). Consider a single person who knows a juicy rumor and wishes to spread it. When a spreader meets an unaware person, the rumor is transmitted and both individuals become spreaders. However, when two spreaders meet, they do not enjoy relating the (now stale) rumor to each other; one or both of them subsequently become stiflers and stop spreading the rumor. The distinction about whether one or both spreaders are converted into stiflers in this interaction is only quantitative. When a spreader and a stifler meet, the spreader is converted into a stifler. All other two-person interactions have no effect.

The densities of the three types of individuals evolve according to

$$\dot{S} = -\alpha S^2 - ST + SU \, ,$$
$$\dot{T} = \quad \alpha S^2 + ST \, , \tag{12.16}$$
$$\dot{U} = -US \, .$$

With $\alpha = 1$, a single spreader converts to a stifler when two spreaders meet (Maki–Thompson model), while $\alpha = 2$ accounts for both spreaders becoming stiflers (Daley–Kendall model). Starting with a single rumor spreader, what is the fraction of the population that remains blissfully unaware of the rumor after it has come and gone by all rumor-mongers having been stifled? For the Maki–Thompson model, this fraction is given by the solution to the transcendental equation $xe^{2(1-x)} = 1$, i.e. $x \doteq 0.203188$ (problem 12.5). For the Daley–Kendall model, the corresponding fraction of clueless individuals is 0.283.

12.2 Discrete reactions

In a continuum description of population dynamics and also chemical kinetics, the basic variables are the global densities of each species. For a finite system, a more complete description is given by the probability that the system contains n_A reactants of type A, n_B of type B, etc. We may think that this distribution sharpens as the system size increases, so that average densities fully characterize the evolution of the reaction. However, this

natural expectation can sometimes be wrong, so that discreteness can play an essential role, even in the thermodynamic limit. We now discuss the effects of discreteness for two general situations: birth–death processes, where particles do not interact, and bimolecular reactions, where an elementary reaction involves two particles.

Birth–death processes

In the *birth process* a system contains identical particles, each of which can give birth to an identical offspring at a constant rate that we set to 1 without loss of generality. How does the number of particles increase with time? In the continuum limit, the average population size $N \equiv \langle n \rangle$ satisfies $\dot{N} = N$ and hence it increases exponentially with time $N = N_0 \, e^t$.

Let us now take into account that the number of particles at time t in a given realization of the reaction is not deterministic but depends on the timing of the individual random birth events. The stochastic description is therefore provided by the set of probabilities $P_n(t)$, where n is the population size. The population distribution $P_n(t)$ evolves according to

$$\dot{P}_n = (n-1)P_{n-1} - nP_n \, . \tag{12.17}$$

The first term on the right accounts for the gain in the probability that n particles exist due to the particle number changing from $n-1$ to n. Since any of the $n-1$ particles can give birth, the rate of the gain process is $(n-1)P_{n-1}$. The loss term has a similar origin.

Suppose that the process begins with one particle: $P_n(0) = \delta_{n,1}$. To solve (12.17) with this initial condition we again apply the exponential ansatz (see pages 139 and 215) and seek a probability distribution in the form $P_n = A \, a^{n-1}$. We can simplify matters still further by noticing that $\sum_{n \geq 1} P_n \equiv 1$, which implies that $A = 1 - a$. Thus using $P_n = (1-a) \, a^{n-1}$, we recast the infinite set of equations (12.17) into a single equation $\dot{a} = 1 - a$, from which $a = 1 - e^{-t}$ and

$$P_n(t) = e^{-t} \left(1 - e^{-t}\right)^{n-1} \, . \tag{12.18}$$

Notice that the average population size

$$\langle n \rangle = \sum_{n \geq 1} nP_n = \frac{1}{1-a} = e^t$$

reproduces the deterministic prediction. However, this agreement does not extend to higher moments of the population distribution. In particular, $\sigma^2 \equiv \langle n^2 \rangle - \langle n \rangle^2 = e^{2t} - e^t$, and therefore $\sigma \simeq \langle n \rangle$. Thus fluctuations are huge and they persist forever. These large fluctuations arise because the population distribution (12.18) does not have a sharp peak at $\langle n \rangle$, but rather is peaked at $n = 1$. Moreover, in the long-time limit, the distribution (12.18) is essentially flat when $n \ll \langle n \rangle$, leading to large fluctuations.

Now suppose that, in addition to each individual giving birth to identical offspring with rate λ, each individual can also die with rate μ. In the continuum limit, the average population size therefore evolves according to $dN/dt = (\lambda - \mu)N$. The solution is $N = N_0 \, e^{(\lambda - \mu)t}$, so that the population grows exponentially in time if $\lambda \neq \mu$, and is static if $\lambda = \mu$. Within the

stochastic framework, the population distribution in this birth–death process is governed by master equations.

$$\dot{P}_n = \lambda(n-1)P_{n-1} - (\lambda+\mu)nP_n + \mu(n+1)P_{n+1}. \tag{12.19}$$

For $n \geq 1$, we again seek a solution in the exponential form $P_n = Aa^{n-1}$. This ansatz immediately leads to[2]

$$A(t) = \mathcal{E}(t)\left(\frac{1-\mu/\lambda}{\mathcal{E}(t)-\mu/\lambda}\right)^2, \qquad a(t) = \frac{\mathcal{E}(t)-1}{\mathcal{E}(t)-\mu/\lambda}, \qquad \mathcal{E}(t) \equiv e^{(\lambda-\mu)t}. \tag{12.20}$$

These formulae have many interesting consequences. Consider first P_0, the probability that there are no particles in the system, that is, the population becomes extinct. This is an *absorbing state* of the discrete dynamics; once the system falls into such an absorbing state, it remains there forever. The existence of absorbing states is one of the primary reasons why the fate of a finite system is usually very different from predictions that follow from continuum descriptions. From Eqs (12.20), we obtain

$$P_0 = \frac{\mu}{\lambda}\frac{\mathcal{E}(t)-1}{\mathcal{E}(t)-\mu/\lambda}. \tag{12.21}$$

In the long-time limit,

$$P_0(t=\infty) = \begin{cases} \dfrac{\mu}{\lambda}, & \lambda > \mu, \\ 1, & \lambda \leq \mu. \end{cases}$$

Thus if the death rate exceeds the birth rate, extinction is sure to occur. Conversely, if the birth rate exceeds the death rate, then the population has a non-zero probability of surviving forever; however, even in this situation the extinction probability is still non-zero. From (12.21) we find that the survival probability $S(t) \equiv 1 - P_0(t)$ decays to its asymptotic value exponentially in time for $\lambda \neq \mu$ and that $S = (1+\lambda t)^{-1}$ for $\lambda = \mu$. Computing the mean population size and the variance we obtain (see problem 12.11)

$$\langle n \rangle = \mathcal{E}(t), \quad \sigma^2 = \langle n^2 \rangle - \langle n \rangle^2 = \frac{\lambda+\mu}{\lambda-\mu}\mathcal{E}(t)[\mathcal{E}(t)-1]. \tag{12.22}$$

The mean population size is again exactly captured by the deterministic continuum approach; however, the deterministic approach remains "morally" wrong; in the interesting case where the birth rate exceeds the death rate, there are again huge fluctuations between different realizations of the process.

Note also that the borderline "static" case of $\lambda = \mu$ is not static at all – it is true that the mean size $\langle n \rangle$ remains constant; however, extinction occurs with probability one and the variance grows with time, $\sigma^2 = 2\lambda t$. This growth of variance arises because extinction

---

[2] Assuming again that the system initially contains a single particle, $P_n(0) = \delta_{n,1}$.

must eventually occur, but $\langle n \rangle$ remains constant. Consequently, in the realizations of the process that do survive for a long time, the population tends to be large.

In hindsight, the inapplicability of the continuum approach is easy to understand. If we start with a single particle, the random nature of individual birth and death events obviously matters. Therefore fluctuations that occur at the very beginning play a huge role. These fluctuations get locked in – as the population increases, the growth becomes progressively more deterministic. To check this idea, consider the birth process and assume that the process begins with $n_0 > 1$ particles, i.e. $P_n(0) = \delta_{n,n_0}$. For $n_0 > 1$, the exponential ansatz fails, but the problem can be solved by more systematic approaches (see problem 12.7). In this case, the probability distribution again has the compact form (problem 12.9)

$$P_n(t) = \binom{n-1}{n_0-1} e^{-n_0 t} \left(1 - e^{-t}\right)^{n-n_0}. \tag{12.23}$$

Using this distribution, the mean population size and the variance are

$$\langle n \rangle = n_0\, e^t\,, \qquad\qquad \sigma^2 = n_0 \left(e^{2t} - e^t\right). \tag{12.24}$$

The relative fluctuations $\sigma/\langle n \rangle = [(1 - e^{-t})/n_0]^{1/2}$ remain finite throughout the evolution. Since their magnitude scales as $(n_0)^{-1/2}$, we conclude that relative fluctuations are indeed small for $n_0 \gg 1$.

Biomolecular reaction: annihilation

The two previous examples described non-interacting particle systems. Interaction usually makes a many-particle problem intractable. Among the few exceptions are single-species annihilation and coalescence, and we now present the former. For single-species annihilation, the density evolves according to the rate equation $d\rho/dt = -\rho^2$, from which $\rho = (1 + t)^{-1}$. The full description is provided again by the set of probabilities $P_n(t)$. The governing equations are

$$\dot{P}_n = \frac{1}{2}\big[(n + 1)(n + 2)P_{n+2} - n(n - 1)P_n\big],$$
$$\equiv \frac{1}{2}\left(\mathbb{E}^2 - 1\right) n(n - 1)P_n. \tag{12.25}$$

For notational convenience we used the raising operator $\mathbb{E}$ defined by $\mathbb{E}f(n) = f(n + 1)$, where f is an arbitrary function of n. Conversely, the inverse, or lowering, operator is defined by $\mathbb{E}^{-1}f(n) = f(n - 1)$. The first term on the right-hand side of Eq. (12.25) accounts for the gain in $P_n(t)$ due to annihilation events in which the particle number $n + 2 \to n$. The rate for such events is proportional to the number of distinct particle pairs in a system of $n + 2$ particles, namely $\frac{1}{2}(n + 1)(n + 2)$. The second term accounts for the loss of particles due to annihilation events in which $n \to n - 2$. To recover the rate equation $d\rho/dt = -\rho^2$, the right-hand side should be multiplied by the inverse of the volume of the system (see problem 12.13 and the following section); here we omit this factor for simplicity.

Before analyzing (12.25) let us pause and think how it is possible that linear equations (12.25) describe an interacting system. In the continuum approximation, interacting systems are usually governed by a *few* coupled nonlinear differential equations. While an exact description of an interacting system can involve linear equations, the catch is that it involves *infinitely* many equations. (An example of how to turn a nonlinear equation into infinitely many linear equations is given in problem 12.12.)

Turning back to Eqs (12.25) we notice that they appear similar to Eqs (12.19), e.g. both sets of equations are linear, so we could hope that Eqs (12.25) are readily tractable. Unfortunately, we cannot exploit the simplicity and power of the exponential ansatz (page 215) because this approach is applicable only when the coefficients are linear functions of n, while the coefficients in (12.25) are quadratic in n.[3] Hence let's use heavier weapons, namely the generating function approach. Introducing the generating function $\mathcal{P}(z,t) = \sum_{n\geq 0} P_n(t)\, z^n$ we recast (12.25) into a partial differential equation,

$$\frac{\partial \mathcal{P}}{\partial t} = \frac{1}{2}(1 - z^2)\,\frac{\partial^2 \mathcal{P}}{\partial z^2}\,. \tag{12.26}$$

The two-body nature of the reaction leads to a second-order, rather than a first-order, partial differential equation for the generating function. This feature makes the solution of bimolecular reactions more complicated than single-particle reactions. The solution to Eq. (12.26) can be obtained by standard methods, but the details are cumbersome and not particularly illuminating. Thus we merely outline the most important features of the method and the solution. The starting point is to seek separable partial solutions, $\mathcal{P}(z,t) = Z(z)T(t)$. Substituting this form into Eq. (12.26), we find $T \sim e^{-\lambda t}$, while Z satisfies the Legendre differential equation

$$Z'' + \frac{2\lambda}{1 - z^2}\,Z = 0\,. \tag{12.27}$$

If the initial number of particles n_0 is even, then the particle number n always remains even. For later convenience, we therefore write $N_0 = 2k_0$ and $n = 2k$. Furthermore, the solution to (12.27) is an even function of z. The normalized even eigenfunctions of (12.27) are

$$Q_{2k}(z) \equiv \sqrt{\frac{k(2k-1)}{4k-1}}\,[P_{2k}(z) - P_{2k-2}(z)]\,,$$

where P_{2k} is the standard Legendre polynomial of order $2k$, and the corresponding eigenvalue is $\lambda = 2k(k-1)$. By matching to the initial condition, $\mathcal{P}(z, t=0) = z^{2k_0}$, and using the orthogonality relations of the Legendre polynomials, the generating function is

$$\mathcal{P}(z,t) = 1 + \sum_{k=1}^{k_0} A_k Q_{2k}(z)\, e^{k(2k-1)t}\,, \tag{12.28}$$

---

[3] One may try to generalize the exponential ansatz to $P_n = A(t)\,\exp[a(t)n + b(t)n^2]$, but since annihilation leads to the decrease of the number of particles, we need a solution in the range $0 \leq n \leq N_0$; due to the additional "boundary" at N_0, it seems impossible that a solution would have a simple form.

with

$$A_k = \sqrt{\frac{4k-1}{k(2k-1)}} \, \frac{\Gamma(k_0+1)\Gamma(k_0+\frac{1}{2})}{\Gamma(k_0+k+\frac{1}{2})\Gamma(k_0-k+1)} \, .$$

From the generating function we can extract moments of the particle number by

$$\langle n^j \rangle = \left(z\frac{\partial}{\partial z} \right)^j \mathcal{P}(z)\big|_{z=1} \, .$$

To compute these derivatives, we use Stirling's approximation to give, for A_k with k large,

$$A_k \simeq \sqrt{\frac{4k-1}{k(2k-1)}} \, e^{-k(2k-1)/2k_0} \, .$$

Moreover, the expressions for the derivatives of the Legendre polynomials at $z = 1$ are[4]

$$P_k'(z=1) = \frac{k(k+1)}{2}, \qquad P_n''(z=1) = \frac{(k-1)k(k+1)(k+2)}{8}, \qquad \text{etc.}$$

Assembling these results and working in the intermediate scaling regime where $1 \ll k \ll k_0$, we obtain

$$\langle n \rangle \simeq \sum_{k=1}^{k_0} 4k \, e^{-2k^2/N}, \qquad\qquad \langle n^2 \rangle \simeq \sum_{k=1}^{k_0} 8k^3 \, e^{-2k^2/N} \, , \qquad\qquad (12.29)$$

where $N = N(0)/(1+\rho(0)t)$ is the time-dependent number of particles from the mean-field solution. We obtain the asymptotic behavior of these sums in the scaling regime by letting the upper limit become infinite and replacing the sums by integrals. This approximation gives

$$\langle n \rangle \simeq N, \qquad\qquad \langle n^2 \rangle \simeq 2N^2 \, ,$$

so that the average number of particles coincides with the mean-field prediction.

Completion time in coalescence

We have thus far focused on the time evolution of the number of reactants. A complementary issue is the time needed for the number of reactants to reach a specified threshold. For example, what is the time for a population to grow to a given size (problem 12.8). Conversely, how long does a population survive before becoming extinct? In annihilation

---

[4] These identities are derived by using the generating function representation for the Legendre polynomials

$$(1 - 2hz + z^2)^{-1/2} = \sum_{k\geq 0} h^k P_k(z),$$

differentiating both sides the requisite number of times, setting $z = 1$, and finally expanding the resulting function on the left-hand side in a power series.

and coalescence, extinction means that a finite system ultimately either becomes empty (annihilation) or ends with a single particle (coalescence). How does the time T to reach this final state – the completion time – depend on the initial number of reactants? What is the distribution of these completion times?

A naive estimate for the completion time can be obtained from the continuum description that gave the density $\rho(t) = 1/(1+t)$. The reaction is complete when one particle remains, corresponding to $\rho(T) \sim 1/N$. This criterion immediately gives $T \sim N$. This is only an estimate; the continuum approach is valid when the number of particles is large, but it is inapplicable at the late stage when only a few particles remain.

Let us now try to find the completion time using an exact approach. The idea is to construct a master equation for this time as a function of the current state of the system; the same idea was previously used in Section 2.6 for computing the exit time for random walks. Let us apply this method to determine the average completion time for coalescence.

The number of reactants at any time is a stochastic variable that changes from n to $n-1$ at a rate $r_n = n(n-1)/N$.[5] The distribution of each individual transition time is then exponentially distributed,

$$\Pi(t_n) = r_n \exp(-r_n t_n),$$

in which the average transition time and its variance are given by $\langle t_n \rangle = r_n^{-1}$ and $\sigma_n^2 = \langle t_n^2 \rangle - \langle t_n \rangle^2 = r_n^{-2}$. Since the average time for the event $n \to n-1$ to occur is $t_n = r_n^{-1}$, the time T_k until k particles remain is

$$T_k = \sum_{n=N}^{k+1} t_n = N \left[\frac{1}{N(N-1)} + \frac{1}{(N-1)(N-2)} + \cdots + \frac{1}{k(k+1)} \right]$$

$$= N \left[\left(\frac{1}{N-1} - \frac{1}{N} \right) + \left(\frac{1}{N-2} - \frac{1}{N-1} \right) + \cdots + \left(\frac{1}{k} - \frac{1}{k+1} \right) \right]$$

$$= N \left[\frac{1}{k} - \frac{1}{N} \right]. \tag{12.30}$$

Thus the average completion time until a single particle remains is $\langle T \rangle \equiv T_1 = N-1$; a similar result arises in annihilation (problem 12.15). Thus the continuum approach gave us a qualitatively correct result.

One can extend the above direct approach to compute the second moment of the completion time. For $N \to \infty$, the result is (problem 12.16)

$$\frac{\langle T^2 \rangle}{\langle T \rangle^2} = \frac{\pi^2}{3} - 2, \tag{12.31}$$

which indicates that the distribution of completion times is non-trivial.

To obtain moments of the completion time in a systematic way, we need the probability distribution $P_n(t)$ that the system contains n particles at time t. Similar to Eq. (12.25), this

---

[5] As discussed previously, the factor $1/N$ is needed to reproduce the mean-field rate equation for the density $d\rho/dt = -\rho^2$ (see also problem 12.13).

probability evolves as

$$\frac{dP_n}{dt} = r_{n+1}P_{n+1} - r_nP_n, \tag{12.32}$$

subject to the initial condition $P_n(0) = \delta_{n,N}$. We impose the boundary conditions $P_{N+1} \equiv 0$, so that this equation applies for all $1 \leq n \leq N$. To solve (12.32), we Laplace-transform it to give

$$(s + r_n)P_n(s) = \delta_{n,N} + r_{n+1}P_{n+1}(s). \tag{12.33}$$

For $n = N$ we get $P_N(s) = (r_N + s)^{-1}$, from which we then solve the rest of Eqs (12.33) recursively. The most relevant quantity for the completion time is $P_1(s)$, the Laplace transform of the probability that the system consists of a single particle. This probability is given by

$$P_1(s) = \prod_{n=2}^{N} \frac{r_n}{s + r_n} = \prod_{n=2}^{N} \left[1 + \frac{sN}{n(n-1)}\right]^{-1}. \tag{12.34}$$

These results hold for any N, but we want to extract the large-N behavior. In this limit the product in (12.34) becomes infinite. To understand what to do with sN we note that the completion time scales as $T \sim N$ and therefore the Laplace variable s should be of the order of N^{-1}. Hence to take the $N \to \infty$ limit we should consider $S \equiv sN$ finite. Thus (12.34) simplifies to

$$P_1(S) = \prod_{n\geq 2} \left[1 + \frac{S}{n(n-1)}\right]^{-1}. \tag{12.35}$$

More fun with gamma functions

Building on the highlight "Fun with gamma functions" on page 156, we now compute

$$P_1(S) = \prod_{n\geq 2} \left[1 + \frac{S}{n(n-1)}\right]^{-1} = \prod_{n\geq 2} \frac{n(n-1)}{(n-n_+)(n-n_-)}, \tag{12.36}$$

with $n_\pm = \frac{1}{2}(1 \pm \sqrt{1-4S})$. As often happens, it is easier to analyze a more general case. Consider the finite product

$$\Pi_N = \prod_{n=2}^{N-1} \frac{(n+a)(n+a^*)}{(n+b)(n+b^*)}. \tag{12.37}$$

We are interested in $\Pi = \lim_{N\to\infty} \Pi_N$, with parameters $a, a^*, b,$ and b^* that satisfy $a + a^* = b + b^*$. Starting with the identity (5.69),

$$(2+a)(3+a)\ldots(N-1+a) = \frac{\Gamma(N+a)}{\Gamma(2+a)},$$

we rewrite (12.37) as

$$\Pi_N = \frac{\Gamma(N+a)\,\Gamma(N+a^*)}{\Gamma(N+b)\,\Gamma(N+b^*)}\,\frac{\Gamma(2+b)\,\Gamma(2+b^*)}{\Gamma(2+a)\,\Gamma(2+a^*)}\,. \tag{12.38}$$

Now using the asymptotic relation given in Eq. (5.71),

$$\frac{\Gamma(N+a)}{\Gamma(N+b)} \simeq N^{a-b} \quad \text{as} \quad N \to \infty,$$

the N-dependent part of (12.38) becomes $N^{a+a^*-b-b^*}$, which equals 1 if $a+a^* = b+b^*$. Thus

$$\Pi = \lim_{N\to\infty} \Pi_N = \frac{\Gamma(2+b)\,\Gamma(2+b^*)}{\Gamma(2+a)\,\Gamma(2+a^*)}\,. \tag{12.39}$$

For $P_1(S)$ in (12.36), $a=0$, $a^*=-1$, $b=-\frac{1}{2}(1-\sqrt{1-4S})$, and $b^*=-\frac{1}{2}(1+\sqrt{1-4S})$, so that $a+a^* = b+b^*$. Specializing (12.39) to these values of a, a^*, b and b^* gives

$$P_1(S) = \Gamma\left(\frac{3-\sqrt{1-4S}}{2}\right)\Gamma\left(\frac{3+\sqrt{1-4S}}{2}\right). \tag{12.40}$$

Finally, we use the identities

$$\Gamma(1+x) = x\Gamma(x), \quad \Gamma\left(\tfrac{1}{2}-x\right)\Gamma\left(\tfrac{1}{2}+x\right) = \frac{\pi}{\cos(\pi x)}\,,$$

to recast (12.40) into the form given in Eq. (12.41).

We now use the results from the highlight (above) to express $P_1(S)$ in the compact form

$$P_1(S) = \frac{\pi S}{\cos\left(\frac{1}{2}\pi\sqrt{1-4S}\right)}\,. \tag{12.41}$$

The moments of the completion time $\langle T^n \rangle$ may now be read off by expanding the Laplace transform (12.41) in a power series of the form $P_1(\sigma) = \sum_{n\geq 0}(-sN)^n\langle T^n\rangle/n!$. The leading behavior of these moments is $\langle T^n \rangle \simeq C_n N^n$, with the first three coefficients $C_1 = 1$, $C_2 = \frac{1}{3}\pi^2 - 2$, and $C_3 = 12 - \pi^2$. Because the Laplace transform (12.41) obeys scaling in the large-N limit, i.e. $P_1(s) \to P_1(S)$ with $S = sN$, the distribution of condensation times is a function of the scaled time T/N,

$$P_1(T) = N^{-1}\,\Phi\,(T/N)\,.$$

To determine the long-time behavior of $P_1(T)$, notice that the Laplace transform $P_1(S)$ has infinitely many poles on the negative real axis at $S_k = -k(k+1)$, with k a positive integer. The closest pole to the origin is at $S = -2$, where $P_1(S) = 6(S+2)^{-1} + \cdots$. This

behavior corresponds to the long-time behavior $P_1(T) \simeq 6\,e^{-2T}$. Similarly, the large-S behavior of $P(S)$ determines the short-time asymptotic of $P_1(T)$. From (12.41), we obtain $P(S) \simeq 2\pi\,Se^{-\pi\sqrt{S}}$ whose inversion can be performed by the Laplace method. From these, the limiting behaviors of the completion time distribution are given by

$$\Phi(x) \simeq \begin{cases} \frac{1}{4}\,(\pi/x)^{7/2}\ e^{-\pi^2/4x}, & x \to 0, \\ 6\,e^{-x}, & x \to \infty. \end{cases} \tag{12.42}$$

We see that the completion time distribution vanishes very quickly away from the peak, so that a completion time that is very different from its average value is extremely unlikely.

12.3 Small-fluctuation expansion

Population dynamics models and chemical reactions are generally more complex than the idealized examples presented above. Consequently, it is not feasible to carry out an exact analysis of their dynamics. If, however, fluctuations of the reactant densities about their averages are small, then a natural way to gain information is through the van Kampen *small-fluctuation expansion* of the master equation. In this approach, the leading term reproduces the rate equation for average densities, and the first correction accounts for fluctuations about these averages. The advantage of this expansion is its generality and systematic nature. The disadvantage is that this expansion only applies when fluctuations are small. As the discussion of the previous section showed, small fluctuations are not as ubiquitous as one might anticipate.

To illustrate the small-fluctuation expansion, let's consider generic single-species reactions in which the fundamental variable is the particle number n. Assuming that fluctuations are small, we write the particle number as

$$n = V\rho(t) + V^{1/2}\xi\,, \tag{12.43}$$

where V is the volume of the system, $\rho(t)$ is the density, and ξ is a stochastic variable. Thus we decompose the particle number into an extensive deterministic component, $V\rho(t)$, and a stochastic component, $V^{1/2}\,\xi$; we have built in the expectation that the magnitude of the stochastic component is of the order of the square root of the extensive component. We also write

$$P_n(t) = \Pi(\xi, t)\,, \tag{12.44}$$

to separate the smoothly varying explicit time dependence and the dependence on the stochastic variable ξ. We are interested in reactions for which the particle number changes by a small amount in each microscopic interaction. These changes are succinctly encoded by the raising and lowering operators $\mathbb{E}$ and $\mathbb{E}^{-1}$, respectively (see Eq. (12.25) and the discussion immediately following). Because the operator $\mathbb{E}$ shifts the particle number by 1,

the variable ξ is shifted by $V^{-1/2}$. As a result

$$\mathbb{E} = 1 + \frac{1}{V^{1/2}} \frac{\partial}{\partial \xi} + \frac{1}{2V} \frac{\partial^2}{\partial \xi^2} + \cdots . \tag{12.45}$$

Finally to convert the time derivative $\dot{P}_n$ in the master equation to the variables (ξ, t), we note that the derivative is taken with n held constant: $dn = V d\rho + V^{1/2} d\xi = 0$, or $d\xi = -V^{1/2} d\rho$. Then the time derivative of P_n becomes

$$\frac{dP_n}{dt} = \frac{d\Pi(\xi, t)}{dt} = \frac{\partial \Pi}{dt} - V^{1/2} \dot{\rho} \frac{\partial \Pi}{\partial \xi} . \tag{12.46}$$

With these preliminaries, the expansion of the master equation for a reacting system becomes an exercise in bookkeeping. As we shall see in the following examples, this expansion naturally divides into extensive terms that describe the evolution of the average density, and sub-extensive terms that describe fluctuations about the average density.

Example 12.5. *Decay $A \to \emptyset$.* The equation for P_n in the decay reaction is

$$\dot{P}_n = -nP_n + (n+1)P_{n+1} = (\mathbb{E} - 1)nP_n . \tag{12.47}$$

Substituting the expressions in Eqs (12.43)–(12.46) into the above master equation, we obtain

$$\frac{\partial \Pi}{\partial t} - V^{1/2} \dot{\rho} \frac{\partial \Pi}{\partial \xi} = \left(\frac{1}{V^{1/2}} \frac{\partial}{\partial \xi} + \frac{1}{2V} \frac{\partial^2}{\partial \xi^2} \right) \left(V\rho + V^{1/2}\xi \right) \Pi. \tag{12.48}$$

Gathering terms order by order gives:

order $V^{1/2}$ $$\qquad\qquad \dot{\rho} \frac{\partial \Pi}{\partial \xi} = -\rho \frac{\partial \Pi}{\partial \xi} , \tag{12.49a}$$

order V^0 $$\qquad\qquad \frac{\partial \Pi}{\partial t} = \frac{\partial (\xi \Pi)}{\partial \xi} + \frac{\rho}{2} \frac{\partial^2 \Pi}{\partial \xi^2} . \tag{12.49b}$$

The first equation recovers the rate equation $\dot{\rho} = -\rho$ with solution $\rho = e^{-t}$; the second is a Fokker–Planck equation with time-dependent coefficients. The general solution of this latter equation is generically a Gaussian function. For our purposes, however, all we need are the moments of the stochastically fluctuating variable $M_k \equiv \langle \xi^k \rangle = \int \xi^k \Pi \, d\xi$. These may be obtained by multiplying Eq. (12.49b) by ξ^k and integrating over all ξ. The integration on the right-hand side can be done straightforwardly by parts. For example, the two terms in the equation for the first moment become

$$\int \xi \frac{\partial (\xi \Pi)}{\partial \xi} \, d\xi = -\int \xi \Pi \, d\xi = -M_1 , \qquad\qquad \int \xi \frac{\rho}{2} \frac{\partial^2 \Pi}{\partial \xi^2} \, d\xi = 0.$$

Similarly in the equation for the second moment,

$$\int \xi^2 \frac{\partial(\xi\Pi)}{\partial\xi}\, d\xi = -\int 2\xi\,(\xi\Pi)\, d\xi = -2M_2\,,$$

$$\int \xi^2 \frac{\rho}{2}\frac{\partial^2\Pi}{\partial\xi^2}\, d\xi = +\int \rho\,\Pi\, d\xi = \rho,$$

in which integration by parts is performed twice in the second computation. In all these partial integrations, the integrated terms are zero because Π vanishes faster than any power law as $\xi \to \pm\infty$. Thus the first two moments evolve in time according to

$$\dot{M}_1 = -M_1\,, \qquad\qquad \dot{M}_2 = -2M_2 + \rho\,.$$

The initial conditions are $M_k(0) = 0$, since the system starts with a deterministic initial condition. Then $M_1(t) = 0$ and we also find $M_2(t) = e^{-t} - e^{-2t}$, which agrees with the exact solution (problem 12.10). Using these results together with (12.43) we get $\langle n \rangle = V\rho$ (since $\langle \xi \rangle = 0$) and

$$\langle n^2 \rangle = V^2\rho^2 + V\langle \xi^2 \rangle = V^2\rho^2 + V\left(e^{-t} - e^{-2t}\right)$$

from which $\sigma^2 = \langle n^2 \rangle - \langle n \rangle^2 = V\left(e^{-t} - e^{-2t}\right)$. The relative fluctuation in the number of particles is therefore

$$\frac{\sigma}{\langle n \rangle} = \sqrt{\frac{e^t - 1}{V}}\,.$$

This fluctuation grows with time and becomes of the order of one when a single particle remains.

Example 12.6. *Annihilation*. We start with the equation $\dot{P}_n = \left(\mathbb{E}^2 - 1\right)n(n-1)P_n/(2V)$, which coincides with (12.25) up to the factor of V^{-1} which was included to recover the proper rate equation for the density. Performing the density expansion, we obtain

$$\frac{\partial\Pi}{\partial t} - V^{1/2}\dot{\rho}\frac{\partial\Pi}{\partial\xi} = \frac{1}{2V}\left(\frac{2}{V^{1/2}}\frac{\partial}{\partial\xi} + \frac{2}{2V}\frac{\partial^2}{\partial\xi^2}\right)\left(V^2\rho^2 + 2V^{3/2}\xi\rho\right)\Pi + \cdots. \quad (12.50)$$

Gathering terms order by order, this gives:

$$\text{order } V^{1/2} \qquad\qquad \dot{\rho}\frac{\partial\Pi}{\partial\xi} = -\rho^2\frac{\partial\Pi}{\partial\xi}\,, \qquad\qquad (12.51a)$$

$$\text{order } V^0 \qquad\qquad \frac{\partial\Pi}{\partial t} = 2\rho\frac{\partial(\xi\Pi)}{\partial\xi} + \rho^2\frac{\partial^2\Pi}{\partial\xi^2}\,. \qquad\qquad (12.51b)$$

Again, from the first equation, we reproduce $\dot{\rho} = -\rho^2$, with $\rho = 1/(1+t)$. Following the same computational steps as in the previous example, the evolution equations for the first two moments are

$$\dot{M}_1 = -2\rho M_1, \qquad\qquad \dot{M}_2 = -4\rho M_2 + 2\rho^2\,,$$

from which $M_1 = 0$ and $M_2 = \frac{2}{3}\left[(1+t)^{-1} - (1+t)^{-4}\right]$, so that the relative fluctuation is

$$\frac{\sigma}{\langle n \rangle} = \sqrt{\frac{M_2}{V\rho^2}} = \sqrt{\frac{2}{3}\frac{1+t-(1+t)^{-2}}{V}}.$$

Again, the relative fluctuation grows with time and becomes of the order of one when a single particle remains.

Example 12.7. *Combined birth–annihilation $A \to A + A$ and $A + A \to \emptyset$. The equation for P_n is now*

$$\dot{P}_n = \frac{1}{2V}\left(\mathbb{E}^2 - 1\right)n(n-1)P_n + \left(\mathbb{E}^{-1} - 1\right)nP_n. \tag{12.52}$$

Applying the density expansion and gathering terms order by order gives:

order $V^{1/2}$ $$\qquad\qquad \dot{\rho}\,\frac{\partial\Pi}{\partial\xi} = (\rho - \rho^2)\,\frac{\partial\Pi}{\partial\xi}, \tag{12.53a}$$

order V^0 $$\qquad\qquad \frac{\partial\Pi}{\partial t} = (2\rho - 1)\,\frac{\partial(\xi\Pi)}{\partial\xi} + \frac{\rho}{2}(1 + 2\rho)\,\frac{\partial^2\Pi}{\partial\xi^2}. \tag{12.53b}$$

As usual, the first equation reproduces $\dot{\rho} = \rho(1 - \rho)$, while the second equation provides the moments. For the second moment we obtain

$$\dot{M}_2 = 2(1 - 2\rho)M_2 + \rho(1 + 2\rho). \tag{12.54}$$

If the system starts in the steady state where $\rho = 1$, then $M_2 = \frac{3}{2}(1 - e^{-2t})$.

The take-home message of the small-fluctuation expansion is that it provides a systematic method to deal with deviations from the average behavior in stochastically interacting particle systems. As long as we are interested in the behavior within a few standard deviations from the average, the statistics of the density fluctuations are Gaussian,

$$\Pi(\xi) = \frac{1}{\sqrt{2\pi\sigma^2}}\,e^{-\xi^2/2\sigma^2}. \tag{12.55}$$

Thus the mean and the variance fully characterize all other moments. Within the framework of the small-fluctuation expansion, the evolution of the system reduces to *deterministic* equations for the first- and second-order moments.

12.4 Large fluctuations

The small-fluctuation expansion discussed above provides a systematic approach for understanding complex reactions. However, a serious pitfall lurks in this expansion in situations

rates

μ_n

λ_n

n

birth > death N death > birth
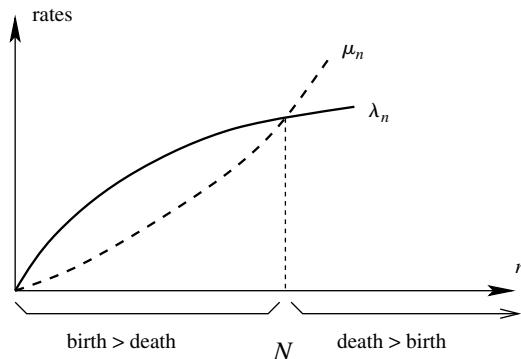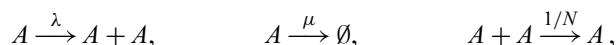
Fig. 12.6.   Generic population dependence of birth and death rates, λ_n and μ_n, so that the continuum system has a steady-state density N, but the stochastic system ultimately reaches $n = 0$.

when the discrete system has an absorbing state. Then a small-fluctuation expansion about the attractor of the rate equation may apply for some time, but will eventually fail and the system will end up in the absorbing state. The birth process discussed above represents one such illustration of this dichotomy. In general, the small-fluctuation expansion fails for state-dependent birth–death processes in which the birth rate exceeds the death rate for small populations, but the death rate exceeds the birth rate otherwise (Fig. 12.6).

Two specific situations where the small-fluctuation expansion fails are birth–death–coalescence and the *SIS* epidemic model. In the former, the elemental reactions are

$$A \xrightarrow{\lambda} A + A, \qquad A \xrightarrow{\mu} \emptyset, \qquad A + A \xrightarrow{1/N} A,$$

while in the *SIS* epidemic model, the reactions between susceptible S and infected I individuals are

$$S + I \xrightarrow{k} I + I, \qquad I \xrightarrow{r} S.$$

For birth–death–coalescence, the rate equation for the density is $d\rho/dt = (\lambda - \mu)\rho - \rho^2$, which gives the fixed point density $\rho^* = \lambda - \mu$ for $\lambda > \mu$. Similarly, for the *SIS* model, the rate equation (already given as Eq. (12.10)) is $dI/dt = (k - r)I - kI^2$. When the intrinsic infection rate k exceeds the recovery rate r, then the fixed point of this equation is $I^* = (k - r)/k$.
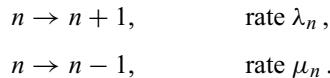
In both these processes, births exceed deaths when the density is less than the fixed point value, while deaths exceed births in the opposite case (Fig. 12.6). Thus the asymptotic state of the corresponding discrete stochastic dynamics is not small fluctuations about the continuum fixed point, but rather extinction! If an unlucky fluctuation leads to extinction, the population cannot recover – a sobering message for ecologically motivated reactions. However, the time until extinction can be exponentially large in the number of particles. One can therefore be easily fooled by numerical simulations if they are not carried out to these exponentially long time scales.

As a warm-up to understanding the extinction time in the class of models embodied by Fig. 12.6, consider first the idealized example of a Brownian particle with diffusivity D in a finite interval $[0, L]$, with a reflecting boundary at $x = L$ and an absorbing boundary at $x = 0$. Suppose that the particle experiences a constant bias velocity v to the right. Most of the time, the particle remains near $x = L$, the bottom of the effective potential well induced by the bias. However, eventually a sufficiently large fluctuation occurs that allows the particle to reach $x = 0$ and subsequently remain trapped there forever. The average time T before the particle becomes trapped scales as (problem 2.20):

$$T \sim \begin{cases} \exp(vL/D), & v > 0, \\ L^2/D, & v = 0, \\ L/v, & v < 0. \end{cases}$$

The huge trapping time for $v > 0$ represents the counterpart of the long extinction time in the stochastic reactions mentioned above.

Let us now investigate the dynamics of extinction for a general birth–death reaction in which the population changes by the following steps:

$$n \to n + 1, \qquad \text{rate } \lambda_n,$$
$$n \to n - 1, \qquad \text{rate } \mu_n.$$

As in Fig. 12.6, the birth and death rates, λ_n and μ_n, satisfy the inequalities $\lambda_n > \mu_n$ for $n < N$ and $\lambda_n < \mu_n$ for $n > N$. To characterize the extinction of the system, we first focus on E_n, the probability of extinction when the system initially has n particles, and T_n, the time until extinction when starting from this same initial state. The extinction probability satisfies the backward equation (see Eq. (2.58))

$$E_n = p_n E_{n+1} + q_n E_{n-1}, \tag{12.56}$$

where $p_n = \lambda_n/(\lambda_n + \mu_n)$ and $q_n = \mu_n/(\lambda_n + \mu_n)$ are the probabilities of making the transitions $n \to n \pm 1$, respectively. The boundary condition for the system of equations (12.56) is $E_0 = 1$.

To solve (12.56), we first rearrange it slightly to read $p_n(E_{n+1} - E_n) = q_n(E_n - E_{n-1})$. Letting $F_n = E_{n+1} - E_n$ and $\rho_n = q_n/p_n = \mu_n/\lambda_n$, we obtain $F_n = \rho_n F_{n-1}$, with solution

$$F_n = F_0 R_n, \qquad R_n = \prod_{j=1}^{n} \rho_j. \tag{12.57}$$

We now sum the F_n to obtain the extinction probability starting from any initial state:

$$E_{n+1} = E_1 + (E_1 - 1) \sum_{i=1}^{n} R_i. \tag{12.58}$$

If the sum $\sum R_i$ is divergent, then necessarily we must have $E_1 = 1$ so that none of the extinction probabilities E_n can exceed 1 or become negative. Concomitantly, we would

obtain $E_n = 1$ for all $n > 1$. Now R_i is an increasing function of i for all $i > N$, since $\mu_i > \lambda_i$ in this range (Fig. 12.6) and therefore $\rho_i = \mu_i/\lambda_i > 1$. The infinite sum $\sum_{i \geq 1} R_i$ therefore diverges. Consequently, we conclude that the extinction probability equals 1 starting with any (finite) number of particles.

Now let's investigate the extinction time. Let T_n denote the time until extinction when the system initially contains n particles. This quantity satisfies the backward equation (see Eqs (2.70)–(2.71))

$$T_n = \frac{\lambda_n}{\lambda_n + \mu_n} T_{n+1} + \frac{\mu_n}{\lambda_n + \mu_n} T_{n-1} + \frac{1}{\lambda_n + \mu_n} , \tag{12.59}$$

subject to the boundary condition $T_0 = 0$. Proceeding in the same way as for the extinction probability, we define $U_n = T_n - T_{n+1}$ to recast (12.59) as the recursion $U_n = (\lambda_n)^{-1} + \rho_n U_{n-1}$. We already know the solution of the homogeneous part of this equation (see (12.57)), so we use it as an integrating factor, that is, we seek U_n in the form $U_n = V_n R_n$. This substitution leads to recursion $V_n = V_{n-1} + (\lambda_n R_n)^{-1}$. We solve for V_n, from which we get U_n and then determine T_n (problem 12.19). The result for general n is

$$T_n = \sum_{j=1}^{n} R_{j-1} \sum_{i=j}^{\infty} \frac{1}{\lambda_i R_i}, \tag{12.60}$$

with agreement that $R_0 = 1$.

Let us focus on the extinction time T_1 when the starting point is just one step away from extinction, because this situation is simple and also natural for many applications. Mathematically, T_1 contains the essential ingredient for general T_n but is a bit simpler:

$$T_1 = \sum_{i \geq 1} \frac{1}{\lambda_i R_i} . \tag{12.61}$$

As a warm-up exercise, consider the rates

$$\lambda_n = n, \qquad \mu_n = \frac{n^2}{N} . \tag{12.62}$$

In this case $\rho_n = n/N$ and $R_n = n!/N^n$, so that

$$T_1 = \sum_{n \geq 1} \frac{N^n}{n \cdot n!} . \tag{12.63}$$

The sum in Eq. (12.63) is similar to the sum in the expansion of the exponential

$$\sum_{n \geq 0} \frac{N^n}{n!} = e^N . \tag{12.64}$$

When $N \gg 1$, the distribution $N^n/n!$ has a sharp peak near $n = N$. The same is true for the distribution $N^n/(n \cdot n!)$. Therefore the terms that provide the dominant contribution to

the sum in (12.63) are N times smaller than the corresponding terms in the sum (12.64). Accordingly, the leading asymptotic behavior of the extinction time is $T_1 \simeq N^{-1}e^N$ when rates are given by (12.62). To compute sub-leading terms, we use the following integral representation of the infinite sum in (12.63),

$$T_1 = \sum_{n \geq 1} \frac{N^n}{n \cdot n!} = \int_0^N \frac{e^X - 1}{X} \, dX \,,$$

and expand near the upper limit to find the asymptotic expansion

$$T_1 = \frac{e^N}{N} \left(1 + \frac{1!}{N} + \frac{2!}{N^2} + \frac{3!}{N^3} + \cdots \right). \tag{12.65}$$

Returning to the general situation we notice that N is usually very large so that the rates λ_n and μ_n can be treated as smooth functions of the continuous variable $x = n/N$. More precisely, we consider the family of rates, which in the large-N limit scale as follows:

$$\lambda_n = Nf(x), \qquad \mu_n = Ng(x), \qquad x = \frac{n}{N}. \tag{12.66}$$

The rates given by Eq. (12.62) belong to this family: namely $f(x) = x$ and $g(x) = x^2$. Another example is represented by the rates

$$\lambda_n = (1 + \epsilon)n - \epsilon \frac{n^2}{N}, \qquad \text{and} \qquad \mu_n = n \tag{12.67}$$

that correspond to the birth–death–coalescence process mentioned at the beginning of this section; in this case $f(x) = x + \epsilon x(1 - x)$ and $g(x) = x$.

We shall see that the function $(\lambda_i R_i)^{-1}$ has a very sharp peak near $i = N$, corresponding to $x = 1$. Using the definition (12.57) for the product variables, taking the logarithm, and replacing the summation by integration, we obtain

$$\ln R_i = \sum_{1 \leq j \leq i} \ln \rho_j = \sum_{1 \leq j \leq i} \ln(\mu_j/\lambda_j)$$

$$\simeq -N \int_0^x \ln \frac{f(y)}{g(y)} \, dy \qquad (x = i/N, \ y = j/N).$$

Using this result and replacing the summation in Eq. (12.61) by integration we obtain

$$T_1 \sim \int_0^\infty \frac{e^{NH(x)}}{f(x)} \, dx \,, \tag{12.68}$$

where we use the shorthand notation

$$H(x) = \int_0^x \ln \frac{f(y)}{g(y)} \, dy \,. \tag{12.69}$$

Our assumptions about the birth and death rates imply that $f(x) > g(x)$ when $x < 1$ and $f(x) < g(x)$ when $x > 1$, which means that $H(x)$ monotonically increases on the interval $0 < x < 1$ and then decreases when $x > 1$. The peak in the integral in (12.68) is very sharp (due to factor $N \gg 1$) and it dominates the integral. Keeping only the leading exponential behavior we find

$$T_1 \sim e^{NH_*}, \qquad H_* \equiv H(1) = \int_0^1 \ln \frac{f(y)}{g(y)} \, dy. \qquad (12.70)$$

For the rates (12.62) we recover $T_1 \sim e^N$; for the rates (12.67) we find $H_* = (1+\epsilon^{-1}) \ln(1+\epsilon) - 1$.

To summarize, when a finite system has an attractive fixed point with respect to the deterministic dynamics, this state is metastable, rather than a true equilibrium when the system also possesses a different absorbing state of the stochastic dynamics. The time to reach this adsorbing state typically scales as e^{NH_*}, with N the population size of the metastable state, and the amplitude H_* is calculable in many situations. Of course, the deterministic behavior can be much more complicated, e.g. a system may have an attractive limiting cycle or exhibit a deterministic chaos. Needless to say, in such situations the extinction time has been computed in a very few cases.

12.5 Notes

Helpful general references that introduce the mathematical techniques for analyzing nonlinear population dynamics equations are given in the texts by Bender and Orszag [138] and by Strogatz [203]. Many of the population biology and epidemic spreading models treated in Section 12.1 are presented in [212]. The model "The struggle for existence" was originally published in 1934 by G. F. Gause [213]. A comprehensive discussion of the three-species competition model (Example 12.3 in this chapter) is given by May and Leonard [214]. A highly readable exposition of the rumor-spreading models is given in [215]; more information about rumor-spreading models can be found in [216].

The coalescence process described in Section 12.2 is quite popular in population biology. In fact, coalescence is essentially the Wright–Fisher model [217, 218] with no selection. If the coalescence process is run backward in time, the coalescence time that was computed in Section 12.4 is the same as the age of the most recent common ancestor. The reader may consult the books [219, 220] for more information about this connection; the articles [221, 222] illustrate the application of tools from physics to the Wright–Fisher model.

The appearance of large fluctuations in single-species birth was pointed out long ago by Delbrück [223]. A useful discussion of the combined birth–death reaction is given by Kendall [224], and this article has extensive references to the earliest literature on this subject. The exact solution for the probability distribution in two-particle annihilation is given by McQuarrie [225, 226]; our approach follows that given in [227]. The small-fluctuation expansion is discussed in detail in van Kampen's book [13]. The use of stochastic methods

for extinction phenomena is a classic field that continues to be vibrant. Two useful books on this subject are [228,229] The approach used to determine the extinction probability and extinction time for a general birth–death process is given in Karlin and Taylor's book [2]. More recent developments on this subject can be found in [230–233].

12.6 Problems

12.1 Generalize the two-species competition model to incorporate competitive asymmetry. The evolution of this reaction is now governed by

$$\dot{A} = A(1 - A - \epsilon_A B) \equiv F_A(A, B),$$
$$\dot{B} = B(1 - B - \epsilon_B A) \equiv F_B(A, B).$$

Study the dynamical behavior of these equations as a function of ϵ_A and ϵ_B and enumerate all possible behaviors.

12.2 Perform a linearized analysis of the Lotka–Volterra equations (12.5) and determine the nature of the local flow near the two fixed points $(0, 0)$ and $(1, 1)$. Next, investigate the following two generalizations of the Lotka–Volterra model.

(a) The constants on the right-hand side of the Lotka–Volterra model are all distinct:
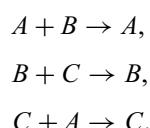
$$\dot{A} = r_A A - k_{AB} AB,$$
$$\dot{B} = -r_B B + k_{BA} AB.$$

(b) The prey species is subject to self-regulation:

$$\dot{A} = A(1 - A) - AB,$$
$$\dot{B} = -B + AB.$$

Determine the fixed points for these two models and describe the global flow in the A–B plane.

12.3 Study the dynamical behavior of the rock–scissors–paper game. Here scissors cuts (beats) paper, rock breaks (beats) scissors, and paper covers (beats) rock. Symbolically, these processes may be represented by

$$A + B \rightarrow A,$$
$$B + C \rightarrow B,$$
$$C + A \rightarrow C.$$

(a) Write the equations that account for the density evolution of the three species.
(b) Determine the evolution of the system when the initial densities are all distinct.

(c) Generalize the rock–scissors–paper game to allow logistic growth to be included for each species. The equations for this generalization are

$$\dot{A} = A(1 - A) - AC,$$
$$\dot{B} = B(1 - B) - BA,$$
$$\dot{C} = C(1 - C) - CB.$$

Determine the dynamical behavior of this system of equations.

12.4 Consider the three-species competition model of Example 12.4 (page 381).

(a) Verify the stability of the coexistence fixed point in the region of parameter space defined by $\alpha + \beta < 2$, and the stability of a single-species fixed point in the region $(\alpha > 1) \cap (\beta > 1)$.

(b) Investigate the dynamical behavior for the case where α and β lie in the complementary region $[\alpha + \beta < 2] \cup [(\alpha > 1) \cap (\beta > 1)]$. Start by writing the evolution equations for the sum, $S = A + B + C$, and the product, $\Pi = ABC$. Show that by neglecting quadratic terms in the equation for $\dot{S}$, the asymptotic solution is $S \to 1$. In this approximation, show that Π asymptotically decays exponentially with time. Thus a trajectory lies in the triangle defined by $A + B + C = 1$, but moves progressively closer to the boundaries of the triangle, since $\Pi \to 0$. Finally, study the trajectory close to the point $(1, 0, 0)$ by assuming that terms quadratic in B and C are vanishingly small and A is close to 1. Use this information to show that the time τ that a trajectory spends in the neighborhood of the point $(1, 0, 0)$ is roughly

$$\tau = \frac{(\alpha + \beta - 2)(\beta - \alpha)t}{2(\beta - 1)(1 - \alpha)}.$$

Thus the time that a trajectory spends near any single-species fixed point scales in proportion to the total elapsed time.

12.5 Use the method of analysis developed to treat the *SIR* model to determine the fraction of a large population that never learns of a rumor that starts with a single spreader.

12.6 Starting with the governing equation (12.17) for the birth process, write the evolution equations for the first three moments of the number distribution $M_j \equiv \langle n^j \rangle$. Solve these equations and show that their time dependences are $M_1 = e^t$, $M_2 = 2e^{2t} - e^t$, and $M_3 = 6e^{3t} - 6e^{2t} + e^t$.

12.7 Instead of using the exponential ansatz, find the probability distribution for the birth process using the generating function technique. Try then to get the solution using induction: solve (12.17) for $n = 1, 2$, and 3 (this can be done thanks to the recurrent nature of the governing equations), then guess P_n in the general case, and verify your guess.

12.8 For the birth process, compute $\langle T_N \rangle$, the average time for which the number of particles reaches N, when the system initially contains a single particle. Also, compute the distribution of times T_N. What is the most probable time at which this probability reaches a maximum?

12.9 Consider the birth process starting with $n_0 > 1$ particles. Derive the results cited in the text, Eqs (12.23)–(12.24).

12.10 Consider the decay process $A \to \emptyset$.

(a) Use the generating function method to solve the master equation and thereby find $P_n(t)$ when initially the system contains $N > 1$ particles.

(b) Find $\langle n \rangle$ and $\langle n^2 \rangle$ and argue that the fluctuations in the number of particles are small.

(c) Starting with N particles, determine the average time until all particles disappear. *Note:* It is possible to write $P_n(t)$ without any calculation. If you can do so, then skip step (a).

12.11 For the birth–death process, compute $\langle n \rangle$ and $\langle n^2 \rangle$ that appear in Eq. (12.22).

12.12 *Linearization.* Let's take a nonlinear ordinary differential equation and turn it into a system of infinitely many linear differential equations. As an example, consider the first-order differential equation $\dot{x} = x^2$, subject to the initial condition $x(0) = 1$.

(a) Let $x_1 \equiv x$, $x_2 \equiv x^2$, and generally $x_j \equiv x^j$. Show that this transformation recasts the original nonlinear differential equation into a system of first-order linear differential equations

$$\dot{x}_j = jx_{j+1}, \quad j = 1, 2, 3, \dots .$$

The initial conditions are $x_j(0) = 1$ for all j.

(b) Try to solve the equations from (a) using linear techniques. If you succeed, you will get $x_j = (1 - t)^{-j}$, which is of course much easier to obtain if you merely solve $\dot{x} = x^2$.

(c) Recall the theorem that a system of linear differential equations with constant coefficients admits a solution on the time interval $0 \le t < \infty$. Can you reconcile this theorem with the outcome of our problem, namely the existence of a solution only on a finite time interval $0 \le t < 1$?

12.13 Starting with Eq. (12.25) for P_n for annihilation, determine the evolution equations for the first three moments of the number distribution $M_j \equiv \langle n^j \rangle$. In particular, show that the correct mean-field rate equation is recovered if the time is rescaled by $1/V$, where V is the volume of the system.

12.14 Write the master equation for the coalescence process $A + A \to A$. Use the generating function method and a suitable variable transformation to show that the generating function for coalescence is the same as Eq. (12.26) for annihilation. Derive the generating function for the initial condition of N particles. Comment on the difference between the generating functions for annihilation and coalescence even though the governing equations for the generating functions are the same.

12.15 Following the same approach as that leading to Eq. (12.30), compute the average completion time to the state where no particles remain for annihilation $A + A \to 0$ when the system starts with N particles.

12.16 Show that the mean-square completion time for coalescence is given by Eq. (12.31). *Hint:* Take into account that for an exponentially distributed random variable x, $\langle x^2 \rangle = 2\langle x \rangle^2$. Hence for an individual transition time $\langle t_n^2 \rangle = 2\langle t_n \rangle^2 = 2r_n^{-2}$.

12.17 Determine the mass of the largest cluster in a finite system in constant-kernel aggregation.

12.18 This is an extended problem to illustrate that the gelation time goes to zero for aggregation in a finite system with homogeneity index $\lambda > 1$. Consider the generalized product kernel $K(i,j) = (ij)^\lambda$, with $\lambda > 1$, for which the master equations are

$$\frac{dc_k}{dt} = \frac{1}{2} \sum_{i+j=k} (ij)^\lambda \, c_i \, c_j - k^\lambda \, c_k \sum_i l^\lambda c_i.$$

Assume that the total number of particles N is large, so that the master equations (with $c_k = N_k/N$, where N_k is the average number of clusters of mass k) are a good approximation.

(a) At short times, neglect the loss terms and show that $c_k \simeq A_k t^{k-1}$, in which the coefficients satisfy the recursion relations

$$(k-1)A_k = \frac{1}{2} \sum_{i+j=k} (ij)^\lambda \, A_i \, A_j$$

for $k \geq 2$ and with $A_1 = 1$.

(b) Determine the dominant contribution to this recursion relation and thereby infer that the average number of clusters of mass k is

$$N_k = Nc_k \sim N[(k-1)!]^{\lambda-1} \, t^{k-1}.$$

(c) Determine the time for the first appearance of a k-mer by invoking the condition $N_k = 1$.

(d) Use the above result for the first-appearance time to estimate the gelation time and show that it is a decreasing function of N for $\lambda > 1$. Thus as $N \to \infty$, there is instantaneous gelation.

12.19 Complete the steps of the derivation for the mean extinction time (12.61) starting from Eq. (12.60).

13 Diffusive reactions

The previous chapter focused on population dynamics models, where the reactants can be viewed as perfectly mixed and the kinetics is characterized only by global densities. In this chapter, we study *diffusion-controlled* reactions, in which molecular diffusion limits the rate at which reactants encounter each other. In this situation, spatial gradients and spatial fluctuations play an essential role in governing the kinetics. As we shall see in this chapter, the spatial dimension plays a crucial role in determining the importance of these heterogeneities.

13.1 Role of the spatial dimension

When the spatial dimension d exceeds a critical dimension d_c, diffusing molecules tend to remain well mixed. This efficient mixing stems from the transience of diffusion in high spatial dimension which means that a molecule is almost as likely to react with a distant neighbor as with a near neighbor. Because of this efficient mixing, spatial fluctuations play a negligible role, ultimately leading to mean-field kinetics. Conversely, when $d < d_c$, nearby reactants react with high probability. This locality causes large-scale heterogeneities to develop, even when the initial state is homogeneous, that invalidate a mean-field description of the kinetics.

To illustrate the role of the spatial dimension in a simple setting, consider the evolution of a gas of identical diffusing particles that undergo either irreversible annihilation or coalescence (see also Section 1.2). Suppose that each particle has radius R and diffusivity D. In a typical reaction time, the change in the concentration c is proportional to the probability that two particles come into contact, while the reaction time is the inverse of the reaction rate. Assuming spatial homogeneity, the meeting probability factorizes into the product of single-particle densities, while the reaction rate for two diffusing particles scales as $k \sim DR^{d-2}$, as implied by dimensional analysis (Section 1.2) and as also derived in Section 2.7. Consequently, the density evolves by[1]

$$\frac{dc}{dt} = -k\,c^2 \,, \tag{13.1}$$

with asymptotic solution $c \simeq (kt)^{-1} \sim (DR^{d-2}\,t)^{-1}$.

---

[1] There is a difference of a factor of 2 in the rate equations of annihilation and coalescence that does not affect the kinetics qualitatively.

For $d > 2$, the reaction rate is an increasing function of particle radius, so that the concentration is a decreasing function of R. These results agree with intuition – larger particles should be more reactive. For $d = 1$, however, the above predictions are clearly incorrect. It is intuitively clear that in one dimension the particle radii cannot affect asymptotic behavior – only the separation between neighboring reactants matters. Hence the concentration depends only on D and t, and then dimensional analysis gives $c \sim (Dt)^{-1/2}$ for $d = 1$. Another way to arrive at this result and also to understand the density decay at the critical dimension is to recall that for $d \leq 2$ the reaction rate acquires the time dependence (see Eq. (2.89) in Section 2.7):

$$k(t) \sim \begin{cases} D^{d/2} t^{(d-2)/2}, & d < 2, \\ D/\ln(Dt/R^2), & d = 2, \\ DR^{d-2}, & d > 2. \end{cases} \tag{13.2}$$

The time dependence of the reaction rate in $d \leq 2$ dimensions stems from the recurrence of diffusion that systematically depletes the reactant density near an absorber. Using this form for $k(t)$ in the rate equation (13.1), the asymptotic behavior of the density becomes

$$c \sim \begin{cases} (Dt)^{-d/2}, & d < 2, \\ (Dt)^{-1} \ln(Dt/R^2), & d = 2, \\ (DR^{d-2}t)^{-1}, & d > 2. \end{cases} \tag{13.3}$$

Hence the density decays more slowly than the mean-field prediction $c \sim t^{-1}$ whenever $d \leq d_c = 2$. Thus the critical dimension is $d_c = 2$ for diffusion-controlled single-species annihilation. This result may seem a bit academic, as the real world is three-dimensional. However, for many diffusion-controlled reactions, the mean-field description is already invalid in three dimensions. Perhaps the most prominent example is diffusion-controlled two-species annihilation, for which $d_c = 4$, as will be discussed in Section 13.3.

The prediction (13.3) is specific to single-species annihilation, but its qualitative features are generic. Namely, for most irreversible reactions there is a critical dimension d_c such that the following hold:

- For $d < d_c$, the mean-field approach is incorrect; the true decay is slower.
- In the marginal case $d = d_c$, the mean-field approach is almost correct. The mean-field approach generally provides the correct, usually a power-law, leading factor, but typically misses a logarithmic prefactor.
- When $d > d_c$, the mean-field approximation gives asymptotically correct predictions.

Thus for a diffusion-controlled irreversible reaction, the density generically exhibits the following behaviors:

$$c \sim \begin{cases} t^{-\beta(d)}, & d < d_c, \\ t^{-\beta_c} (\ln t)^{\gamma}, & d = d_c, \\ t^{-\beta_c}, & d > d_c, \end{cases} \tag{13.4}$$

with $\beta(d) < \beta_c$ when $d < d_c$ and $\beta_c \equiv \beta(d_c)$. The exponent $\beta(d)$ is not merely smaller than β_c, it is usually a strictly increasing function of d on the interval $d < d_c$. This dependence is exemplified by the behavior of the exponent $\beta(d)$ in single-species and two-species annihilation, where $\beta(d) = d/2$ and $\beta(d) = d/4$, respectively. At the critical dimension, the exponent γ is non-negative because the mean-field description is either slower by a logarithmic factor (as in single-species annihilation at $d_c = 2$, where $\gamma = 1$) or correct (as in two-species annihilation at $d_c = 4$, where $\gamma = 0$).

The values $d_c = 2$ and $d_c = 4$ for the critical dimension of single- and two-species annihilation are common. Anomalous behaviors occur at these four dimensions in equilibrium critical phenomena and in the random walk. For example, the recurrence of random walks for $d \leq 2$ and their transience for $d > 2$ (Section 2.5) essentially determines the critical dimension for the single-species annihilation. A critical dimension $d_c = 4$ characterizes a number of interacting diffusive systems, such as a self-avoiding random walk.

Example 13.1. *Self-avoiding random walk.* In a discrete-time self-avoiding walk on a lattice, a walker hops to a random neighboring site if this site has not been visited previously. For a proper description of the self-avoiding walk we must keep track of its entire trajectory. This infinite memory makes the self-avoiding walk much more complicated than the simple random walk. Nevertheless, the asymptotic behavior of the self-avoiding walk is largely understood, except for the three-dimensional case where the exponents are known only numerically.[2] After $t \gg 1$ time steps, the displacement of the self-avoiding walk averaged over all possible paths grows as

$$
R \sim \begin{cases}
t, & d = 1, \\
t^{3/4}, & d = 2, \\
t^{\nu} \quad \text{with} \quad \nu \approx 0.59, & d = 3, \\
t^{1/2}(\ln t)^{1/8}, & d = 4, \\
t^{1/2}, & d > 4.
\end{cases}
$$

Therefore the critical dimension is $d_c = 4$.

Are there situations for which the critical dimension is different from two and four? Here are two examples where the critical dimension attains the extreme values $d_c = 1$ and $d_c = \infty$, respectively.

Example 13.2. *Three-particle reactions.* Consider the single-species diffusion-controlled process in which a reaction occurs only when three particles come into contact; for example, three-particle annihilation $A+A+A \to \emptyset$ or three-particle coalescence $A+A+A \to A$. While three-particle processes may seem artificial we have seen that they arise in cluster dynamics (Section 8.8) and in extremal dynamics (Section 9.7). In addition, most biological reactions between two molecules require the presence of a catalyst. These are not single-species processes, but they do involve at least three molecules.

---

[2] An excellent approximation is the Flory formula $\nu = 3/(d+2)$ which is exact in one dimension, happens to be exact in two dimensions, and becomes exact at the critical dimension.

We now use rate equations and dimensional analysis to argue that $d_c = 1$. As in the binary reaction, we write the (mean-field) rate equation

$$\frac{dc}{dt} = -k\,c^3\,,$$

whose asymptotic solution is $c \simeq (2kt)^{-1/2}$. Next we use dimensional analysis to express the reaction rate via the radius[3] and the diffusion constant: $k \sim DR^{2d-2}$. Therefore

$$c \simeq (2kt)^{-1/2} \sim \frac{1}{R^{d-1}}\frac{1}{\sqrt{Dt}}\,.$$

The R dependence is reasonable when $d > 1$, while for $d < 1$ it is obviously wrong. Thus we conclude that $d_c = 1$. It turns out that a logarithmic correction is present at the critical dimension:

$$c \sim \sqrt{\frac{\ln(Dt/R^2)}{Dt}} \qquad \text{when} \quad d = 1.$$

This example demonstrates that it is useful to think about dimension d as a real number, even though physically it must be one of three integers: $d = 1, 2$ or 3. Indeed, we are able to predict $d_c = 1$ by extending the spatial dimension to the unphysical range $d < 1$. To have a reaction–diffusion process with $d_c = \infty$ (i.e. the mean-field prediction is incorrect in all dimensions) we must ensure that there is no mixing. Here is a simple example.

Example 13.3. *The trapping reaction.* Consider a medium that contains immobile traps and diffusing particles that are non-interacting among themselves. Whenever a particle touches a trap, it gets absorbed. It is natural to assume sizeless particles, and spherical traps of the same radius R that are randomly distributed in the system at concentration ρ. The lack of interactions between particles allows us to deal with a single-particle problem. At first sight the problem seems simple – if we know the locations of all traps and the initial location of the particle, we must solve the diffusion equation with an absorbing boundary condition on the surfaces of all traps, and then average over all initial positions to determine the survival probability $\mathcal{S}(t)$. This boundary is extremely complicated, however, and it is impossible to extract meaningful results using such a straightforward approach.

A mean-field description is based on the rate equation

$$\frac{d\mathcal{S}}{dt} = -k\,\rho\,\mathcal{S}\,, \tag{13.5}$$

which leads to an exponential decay of the density

$$\mathcal{S} = e^{-k\rho t}\,. \tag{13.6}$$

---

[3] We define the three-body reaction process in an obvious way: Whenever the centers of three particles come within a distance $\le R$ of each other, they react. In a lattice formulation, a reaction occurs whenever three particles occupy the same lattice site; in this case, the lattice spacing plays the role of R.

In the next section we will see that this prediction is not valid for all dimensions; the correct asymptotic is

$$\mathcal{S} \sim \exp\left[-\text{const.} \times \rho^{2/(d+2)} (Dt)^{d/(d+2)}\right]. \tag{13.7}$$

The mean-field prediction is asymptotically incorrect because traps are immobile, so that the disorder in their initial location is quenched forever and mixing never occurs.

Another example of a reaction–diffusion process with $d_c = \infty$ is the following island growth model that we studied in Section 5.5.

Example 13.4. *Island growth model.* Here, each cluster occupies a single lattice site. Monomers are mobile while islands (clusters of mass ≥ 2) are immobile. Symbolically,

$$M + M \to I, \qquad M + I \to I.$$

On the mean-field level, the governing rate equations

$$\frac{dM}{dt} = -2M^2 - MI, \qquad \frac{dI}{dt} = MI,$$

predict an exponential decay of the monomer density. (This computation was the subject of problem 5.5.) The island growth model resembles the trapping reaction, except that the islands (which play the role of traps) are dynamically generated; if initially there are no islands, they will arise during the evolution. The spatial distribution of islands is generated by this process, rather than an initial condition (as in the trapping reaction), and the exact decay of the monomer density has not been computed. This decay certainly depends on the spatial dimension and should resemble (13.7) that characterizes the trapping reaction. These facts suggest that the critical dimension must be infinite, $d_c = \infty$.

Finally, we show below how to artificially tune the spatial dimension in a diffusion-controlled reaction.

Example 13.5. *Non-integer dimensions.* Consider a system of non-interacting passive particles that diffuse and are driven by a velocity field $\mathbf{V} \equiv \mathbf{V}(\mathbf{r}, t)$. The particle density $c \equiv c(\mathbf{r}, t)$ is described by the convection–diffusion equation

$$\frac{\partial c}{\partial t} + (\mathbf{V} \cdot \nabla)\, c = D\nabla^2 c.$$

For a radial velocity field, $\mathbf{V} = V(r)\, \mathbf{r}/r$, and radially symmetric initial conditions, the convection–diffusion equation for the density $c = c(r, t)$ reduces to

$$\frac{\partial c}{\partial t} + V\frac{\partial c}{\partial r} = D\left(\frac{\partial^2 c}{\partial r^2} + \frac{d-1}{r}\frac{\partial c}{\partial r}\right).$$

Imagine now that the radial velocity field is inversely proportional to the distance from the origin, $V = Q/(2\pi r)$. In this case, the convection term can be absorbed into the diffusion

term by shifting the dimension. Thus we can interpret the convection–diffusion process as pure diffusion in a space of non-integer dimension. The radial velocity field $V = Q/(2\pi r)$ is natural in two dimensions where it obeys the continuity equation, $\nabla \cdot \mathbf{V} = 0$, and corresponds to the flow induced by the point source of strength Q. Hence let's consider the two-dimensional case, $d_{\mathrm{phys}} = 2$, with velocity field $V = Q/(2\pi r)$. This convection–diffusion process in two spatial dimensions is mathematically described by the diffusion equation in a fictitious dimension d_{fict} that is not necessarily an integer:

$$\frac{\partial c}{\partial t} = D\left(\frac{\partial^2 c}{\partial r^2} + \frac{d_{\mathrm{fict}} - 1}{r}\frac{\partial c}{\partial r}\right), \qquad d_{\mathrm{fict}} = 2 - \frac{Q}{2\pi D}.$$

For source flows ($Q > 0$) the fictitious dimension is less than the physical dimension, $d_{\mathrm{fict}} < 2$, and vice versa for sink flows ($Q < 0$).

13.2 The trapping reaction

In the trapping reaction, particles diffuse and are absorbed whenever they come into contact with traps that are randomly distributed throughout the system. What is the probability that a particle "survives" until time t? A naive way to study this question is to replace a configuration of traps by an effective average medium. Such a replacement (see Example 13.2 in the previous section) leads to a survival probability that decays exponentially in time, Eq. (13.6). This prediction is wrong. We may try to improve upon this description by recalling that the reaction rate, which was treated as constant in solving (13.5), actually acquires a time dependence when $d \leq 2$. Solving (13.5) with $k(t)$ given by (13.3) we arrive at asymptotics

$$\ln \mathcal{S} \sim -\rho \times \begin{cases} (Dt)^{d/2}, & d < 2, \\ Dt/\ln(Dt/R^2), & d = 2, \\ DR^{d-2}t, & d > 2, \end{cases} \tag{13.8}$$

apparently implying that $d_c = 2$. The effective-medium approximation (13.8) is closer to the true asymptotic, yet it is still incorrect. The predictions (13.6) and (13.8) are erroneous for interesting reasons. As we shall discuss, extreme fluctuations in the spatial distribution of traps lead to a slower-than-exponential decay of the survival probability. This anomalously slow decay sets in only at very long times, at which point the density is vanishingly small. Hence even though extreme fluctuations are crucial from a theoretical perspective, their practical consequences are limited.

Exact solution in one dimension

The essence of the trapping reaction can be best appreciated in one dimension. To obtain the *average* survival probability, we first need to determine the survival probability of a

diffusing particle in a specified configuration of traps, then average over all starting points of the walk, and finally average over all trap configurations. In one dimension, this calculation is not difficult because a diffusing particle "sees" only a single confining interval defined by the nearest surrounding traps. Suppose that the particle is at x_0 and that the nearest traps are at $x = 0$ and $x = L$. The exact concentration at time $t > 0$ for this configuration is given by the Fourier series (problem 13.1)

$$c_L(x, t|x_0) = \frac{2}{L} \sum_{n \geq 1} \sin\left(\frac{n\pi x}{L}\right) \sin\left(\frac{n\pi x_0}{L}\right) e^{-(n\pi)^2 Dt/L^2}. \tag{13.9}$$

We now compute the survival probability in the interval $[0, L]$ by averaging over all initial particle positions and integrating over all final positions within the interval to give

$$\overline{S_L(t)} = \frac{1}{L} \int_0^L \int_0^L c_L(x, t|x_0) \, dx \, dx_0$$

$$= \frac{8}{\pi^2} \sum_{m \geq 0} \frac{1}{(2m+1)^2} e^{-(2m+1)^2 \pi^2 Dt/L^2}, \tag{13.10}$$

where the overbar denotes the average over all starting positions. Finally, we obtain the configuration-averaged survival probability by averaging $\overline{S_L(t)}$ over the distribution of lengths of trap-free intervals. If the traps are randomly distributed at density ρ, the probability that the interval $[0, L]$ is free of traps and that there is a trap between L and $L + dL$ is $P(L) = e^{-\rho L} \rho \, dL$. Thus the formal solution for the average survival probability is the double average over all starting points of the walk and over all trap-free regions:

$$S(\tau) \equiv \langle \overline{S_L(t)} \rangle = \frac{8}{\pi^2} \int_0^\infty d\ell \sum_{m \geq 0} \frac{1}{(2m+1)^2} \exp\left[-\frac{(2m+1)^2 \tau}{2\ell^2} - \ell\right], \tag{13.11}$$

where the angle brackets denote the average over all interval lengths and instead of t and L we use the dimensionless variables

$$\tau = 2\pi^2 \rho^2 Dt, \qquad \ell = \rho L.$$

The time dependence of S is subtle, with different behaviors in the long- and short-time limits. Let's first focus on the long-time behavior of S, where the first term in the above series dominates. Keeping only this leading term, we obtain

$$S(\tau) \simeq \frac{8}{\pi^2} \int_0^\infty d\ell \, e^{f(\ell)}, \quad \text{with} \quad f(\ell) = -\frac{\tau}{2\ell^2} - \ell.$$

The function $f(\ell)$ has a peak at $\ell_* = \tau^{1/3}$ that becomes progressively sharper as $\tau \to \infty$. We may therefore determine the long-time behavior of S by the Laplace method, and the final result is (problem 13.2)

$$S(\tau) \simeq \frac{8}{\pi^2} \sqrt{\frac{2\pi}{3}} \, \tau^{1/6} \exp\left(-\frac{3}{2} \tau^{1/3}\right), \qquad \tau \to \infty. \tag{13.12}$$

This surprising stretched exponential asymptotic decay stems from the competition between the exponentially small probability to find a large trap-free region and the long lifetime of a particle in such a region. Optimizing with respect to these two features is equivalent to maximizing $f(\ell)$. This maximization predicts that the length ℓ^* of the optimal trap-free interval scales as $\tau^{1/3}$.

For the short-time behavior of the survival probability, it is simpler to first consider the derivative of Eq. (13.11) with respect to τ:

$$\frac{d\mathcal{S}}{d\tau} = -\frac{8}{\pi^2} \int_0^\infty d\ell \, \frac{e^{-\ell}}{2\ell^2} \sum_{m \geq 0} \exp\left[-\frac{(2m+1)^2 \tau}{2\ell^2}\right]. \qquad (13.13)$$

When $\tau \to 0$, we can replace the summation by integration:

$$\sum_{m \geq 0} \exp\left[-\frac{(2m+1)^2 \tau}{2\ell^2}\right] \simeq \int_0^\infty dm \, \exp\left[-\frac{2\tau}{\ell^2} m^2\right] = \frac{1}{2} \sqrt{\frac{\pi \ell^2}{2\tau}}. \qquad (13.14)$$

Therefore Eq. (13.13) becomes

$$\frac{d\mathcal{S}}{d\tau} \simeq -\frac{2}{\pi^2} \sqrt{\frac{\pi}{2\tau}} \int_{\sqrt{\tau}}^\infty d\ell \, \frac{e^{-\ell}}{\ell}. \qquad (13.15)$$

The lower limit of the integral is now $\ell = \sqrt{\tau}$ rather than $\ell = 0$ because the replacement of the sum by integration applies only when $\tau \ll \ell^2$. The leading behavior of the integral in (13.15) is $\ln(\tau^{-1/2})$ and therefore $d\mathcal{S}/d\tau \simeq -(2\pi^3\tau)^{-1/2} \ln(1/\tau)$, from which

$$\mathcal{S}(\tau) \simeq 1 - \sqrt{\frac{2\tau}{\pi^3}} \, \ln(1/\tau) \quad \text{when} \quad \tau \to 0. \qquad (13.16)$$

The crossover between short- and long-time behaviors is gradual. Consequently, it is difficult to resolve the long-time asymptotic behavior from experimental data or by simulations. The asymptotic survival probability of $\exp(-t^{1/3})$ can only be observed by clever simulation techniques that are specifically tailored to probe the regime where the survival probability becomes truly minuscule.

General spatial dimension

We can extend the optimal fluctuation argument from one dimension to arbitrary spatial dimension to give a lower bound for the asymptotic survival probability. We shall argue that this lower bound is asymptotically exact in the limit of a small concentration of traps. Without loss of generality, consider a particle that starts at the origin. For a specific configuration of traps, the survival probability is clearly bounded from below by ignoring the traps and placing an absorbing sphere that is centered about the origin whose radius equals the distance to the closest trap (Fig. 13.1(a)). For this inscribed sphere, it is straightforward to calculate the survival probability. We must solve the diffusion equation, subject to the boundary
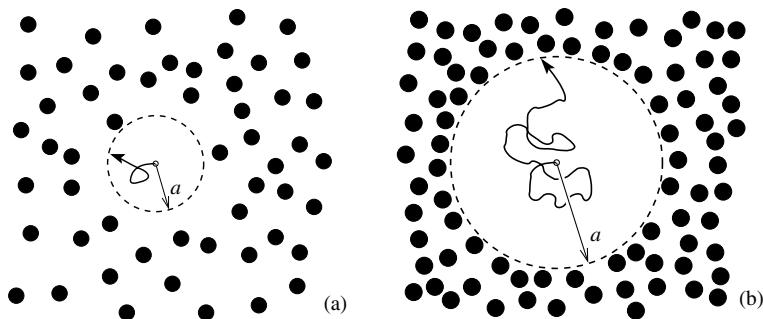
Fig. 13.1. (a) A configuration of traps (circles) and a diffusing trajectory that remains within an inscribed trap-free spherical region of radius a that is centered at the initial particle position. (b) As the radius a becomes larger the trap-free region becomes progressively more spherical.

condition $c(|\mathbf{r}| = a, t) = 0$ and the initial condition $c(\mathbf{r}, t = 0) = \delta(\mathbf{r})$. By separation of variables, the solution can be written as the eigenfunction expansion (problem 13.3)

$$c(r, t) = \sum_{n \geq 1} A_n x^\nu J_\nu(\lambda_n x) \, e^{-\lambda_n^2 Dt/a^2}, \qquad x = \frac{r}{a}. \tag{13.17}$$

Here J_ν is the Bessel function of order $\nu = (2 - d)/2$, λ_n is the nth zero of this Bessel function, and the amplitudes A_n are determined by the initial condition.

We now average over all configuration of traps. This procedure is equivalent to averaging over the radius of the trap-free spherical region. The probability for a trap-free ball of radius r to exist is $e^{-\rho V}$, where $V = V_d r^d$ is the volume of a d-dimensional ball of radius r. As in one dimension, the first term in this series (13.17) dominates in the long-time limit and the lower bound for the survival probability is therefore

$$\mathcal{S}(t) \gtrsim \int_0^\infty \exp\left(-\lambda_1^2 Dt/a^2 - \rho V_d \, a^d\right) \rho \, \Omega_d \, a^{d-1} \, da, \tag{13.18}$$

with Ω_d the surface area of the unit d-dimensional ball.

We now determine the asymptotic behavior of this integral by the Laplace method. The details are parallel to those given for the one-dimensional case and the final result is

$$\ln \mathcal{S} \simeq -C_d \, \rho^{2/(d+2)} \, (Dt)^{d/(d+2)}, \quad C_d = \frac{d+2}{d} \left(\frac{\lambda_1^d \Omega_d}{2}\right)^{2/(d+2)}. \tag{13.19}$$

Therefore the asymptotic behavior depends only on the smallest eigenvalue λ_1. In two dimensions, λ_1 is the smallest positive root of the Bessel function J_0; numerically $\lambda_1 \doteq 2.4048$. In three dimensions, $\lambda_1 = \pi/2$ and the asymptotic (13.19) becomes

$$\ln \mathcal{S} \simeq -\frac{5}{3} \left(\frac{\pi^2}{2}\right)^{4/5} \rho^{2/5} \, (Dt)^{3/5}.$$

As in the case of one dimension, this stretched exponential decay is not obtainable by traditional approaches, such as a perturbative expansion in the limit of a small trap density, and the asymptotic survival probability is not detectable until the density has decayed to a vanishingly small value.

Finally, let us determine the quality of the lower bound (13.19); remarkably, (13.19) is an exact asymptotic.[4] This property is far from evident – keeping only the dominant term in the eigenfunction expansion does not diminish the asymptotic, while placing an artificial spherical absorbing boundary whose radius equals the distance to the closest trap certainly decreases the survival probability. Let us compare the actual void with the largest inscribed sphere. In the long-time limit, when the voids that provide the dominant contribution are also large, they clearly become more and more spherical as suggested by Fig. 13.1(b). This feature is intuitively clear, but hard to prove. Hence we give another argument for the optimality of the sphere.

The best chance for a particle to survive is to be in a large void. The competition between increasing the survival probability by being in such a void and decreasing the existence probability of a void by increasing its size selects the optimal void. In higher dimensions the voids differ not only in their volume but also in their different shapes. Inside an arbitrary void $\mathcal{V}$ of large volume $V = \text{vol}(\mathcal{V})$ (that is, $\rho V \gg 1$), the solution of the diffusion equation with an absorbing boundary condition can be written as the eigenfunction expansion

$$c(r, t) = \sum_{n \geq 1} A_n \Phi_n(\mathbf{r}) \, e^{-\Lambda_n^2 Dt} \, , \tag{13.20}$$

where the eigenvalues and eigenfunctions are determined by the Laplace operator with Dirichlet (absorbing) boundary condition,

$$\nabla^2 \Phi_n = -\Lambda_n^2 \Phi_n \, , \quad \Phi_n \big|_{\partial \mathcal{V}} = 0.$$

Dimensional analysis tells us that $\Lambda_n = V^{-1/d} \lambda_n$, with dimensionless eigenvalues λ_n that do not depend on the volume of the void. Repeating a now standard argument we write the long-time asymptotic as

$$\mathcal{S} \sim \int \exp\left(-\lambda_1^2 Dt/V^{2/d} - \rho V\right) \rho \, dV$$

and find

$$\ln \mathcal{S} \simeq -B_d \rho^{2/(d+2)} \left(\lambda_1^2 Dt\right)^{d/(d+2)} \, , \qquad B_d = \left(\frac{d}{2}\right)^{2/(d+2)} + \left(\frac{2}{d}\right)^{d/(d+2)} . \tag{13.21}$$
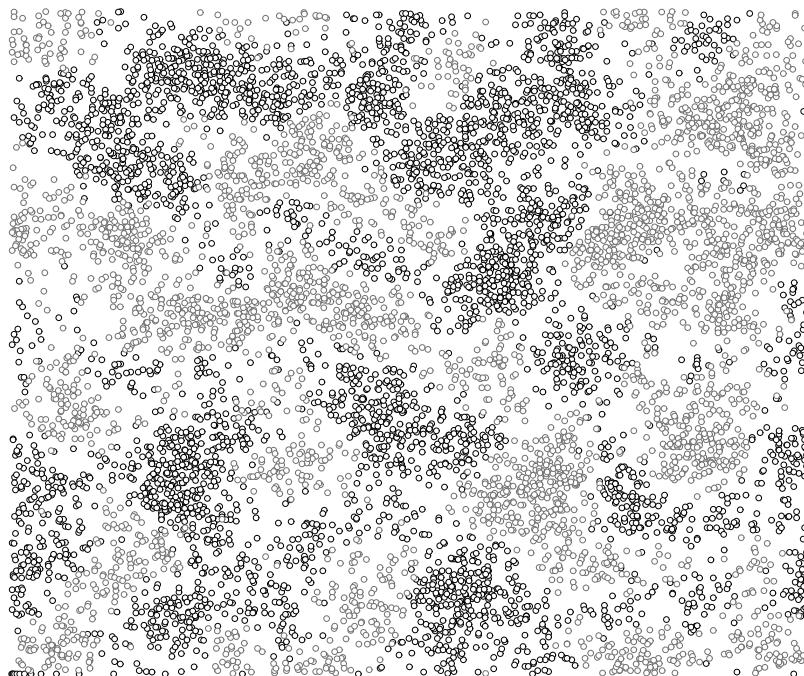
The slowest decay (which dominates the asymptotic) corresponds to voids with minimal first eigenvalue λ_1. Such voids are known to be spherical.[5]

---

[4] In the limit where the volume fraction occupied by traps is negligibly small, $\rho R^d \to 0$.
[5] This is ensured by the Faber–Krahn theorem.

13.3 Two-species annihilation

In *two-species annihilation*, diffusing complementary species A and B annihilate, while same-species particles do not interact. Physical examples of this reaction include electron–hole recombination in a semiconductor, or matter–antimatter annihilation in a cosmological setting. Perhaps the most striking aspect of diffusion-controlled two-species annihilation is that the density decays extremely slowly with time – as $t^{-d/4}$ for spatial dimension $d < 4$ when the initial densities of the two species are equal. This decay is slower than the rate equation prediction of t^{-1}, and also slower than the $t^{-d/2}$ decay of single-species reactions for $d < 2$. Anomalously slow kinetics arises because opposite-species reactants organize into a coarsening domain mosaic (Fig. 13.2). As a result, reactions can occur only along domain boundaries, rather than uniformly throughout the system. This restriction of the reaction to a small subset of the entire system underlies the slow density decay. The $t^{-d/4}$ density decay has been proved by making the heuristic arguments of Section 1.3 in rigorous statements. In spite of what we have learned from both the heuristic and rigorous approaches, the amplitudes of the asymptotic behavior and the full functional form of the density are unknown and remain a formidable challenge, even in one dimension.

Fig. 13.2.   Snapshot of particle positions in two-species annihilation in two dimensions. By squinting at this figure, the three independent length scales should become apparent: the small spacing between neighboring same-species particles, the characteristic domain size, and the width of the interface between opposite-species domains.

Diffusive reactions
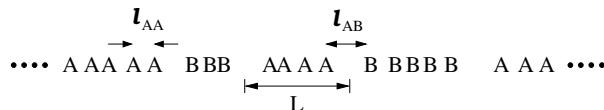
Fig. 13.3. Schematic illustration of the three basic length scales in two-species annihilation in one dimension: the domain size $L \sim t^{1/2}$, the interparticle spacing, which scales as $\ell_{AA} = \ell_{BB} \sim t^{1/4}$, and the depletion zone between domains, which scales as $\ell_{AB} \sim t^{3/8}$ in one dimension.

A basic consequence of the $t^{-d/4}$ density decay is that, below the critical dimension, the average distance between neighboring particles grows in a universal (that is, dimension-independent) manner: $\ell \sim c^{-1/d} \sim t^{1/4}$. One might guess that the length scale ℓ uniquely characterizes the coarsening mosaic of domains in two-species annihilation. This anticipation is erroneous, however, as is evident from Fig. 13.2. In the rest of this section, we use heuristic approaches to argue that the coarsening mosaic in two-species annihilation is characterized by at least three length scales.

The existence of three scales is most easily appreciated in one dimension, where the system organizes itself into alternating A and B domains in the long-time limit (Fig. 13.3). The typical length of a domain is

$$L \sim \sqrt{Dt} \, . \tag{13.22}$$

This scale arises because diffusive transport ensures that there is efficient mixing on a scale $L \sim \sqrt{Dt}$ at time t, so that the species which was in a local minority in such a domain will be eliminated. Since the average initial number of particles of each species on this spatial scale is $c(0)L$ and fluctuations are of the order of $\sqrt{c(0)L}$, the density within a domain at time t is given by

$$c \sim \frac{\sqrt{c(0)L}}{L} \sim \sqrt{c(0)} \, (Dt)^{-1/4} \, , \tag{13.23}$$

as presented in Section 1.3. Using (13.23) and the relation $\ell \sim c^{-1}$, the typical distance between neighboring, same-species particles is

$$\ell_{AA} = \ell_{BB} \sim [c(0)]^{-1/2} (Dt)^{1/4} \, . \tag{13.24}$$

Thus there are at least two scales: the interparticle separation $\ell_{AA} = \ell_{BB}$ and the domain size L. What is a bit surprising is that there is also a third independent scale ℓ_{AB} that characterizes the separation between neighboring opposite-species particles, or equivalently the separation between neighboring domains (Fig. 13.3). To determine ℓ_{AB}, notice that a typical AB pair reacts in a time $\Delta t \sim \ell_{AB}^2/D$. The number of reactions per unit length is of the order of one reaction per domain; that is, $(Dt)^{-1/2}$. Consequently, the rate of change of

the concentration is given by

$$\frac{dc}{dt} \sim \frac{\Delta c}{\Delta t} \sim -\frac{(Dt)^{-1/2}}{\ell_{AB}^2/D} \ . \tag{13.25}$$

However dc/dt is known independently from Eq. (13.23). Using this result for dc/dt, together with (13.25), we obtain

$$\ell_{AB} \sim [c(0)]^{-1/4} (Dt)^{3/8} \ . \tag{13.26}$$

Asymptotically $\ell_{AB} \gg \ell_{AA} = \ell_{BB}$, but $\ell_{AB} \ll L$. The feature that $\ell_{AB} \gg \ell_{AA}$ can be viewed as a manifestation of an effective repulsion between opposite-species particles, as evidenced by the readily visible blank space between opposite-species domains in Fig. 13.2.

We can extend this line of reasoning to two dimensions. Again, in a time scale of the order of $\Delta t \sim \ell_{AB}^2/D$, all particles on the domain periphery will interact. The number of such particles is of the order of the domain perimeter divided by the spacing between periphery particles. A subtle point is that the spacing between these particles is ℓ_{AB} and not ℓ_{AA}. Consequently, the concentration of periphery particles is of the order of $[(Dt)^{1/2}/\ell_{AB}]/Dt$. With these ingredients, the counterpart of Eq. (13.25) becomes

$$\frac{dc}{dt} \sim \frac{\Delta c}{\Delta t} \sim -\frac{[(Dt)^{1/2}/\ell_{AB}]/Dt}{\ell_{AB}^2/D} \ . \tag{13.27}$$

Again using the known result for $c(t)$ itself from Eq. (13.23), we obtain

$$\ell_{AB} \sim [c(0)]^{-1/6} (Dt)^{1/3} \ . \tag{13.28}$$

Above two dimensions, random walks are transient and we can no longer be assured that all periphery particles will react in a time of the order of ℓ_{AB}^2/D. Both an extension of the heuristic arguments presented here and numerical evidence indicate that the anomalous scaling of AB distances disappears in three dimensions so that $\ell_{AB} \sim \ell_{AA} \sim t^{1/4}$.

The spatial organization of reactants into single-species domains breaks down for $d > 4$ because such domains become transparent to an intruder of the opposite species. Consider, for example, the fate of an A particle that is placed at the center of a B domain of linear dimension L. For such a domain, the local concentration of Bs is, from Eq. (13.23), of the order of $L^{-d/2}$. The A intruder needs L^2 time steps to exit the domain, during which time L^2 distinct sites will be visited (again assuming $d > 4$). At each site, the A particle will react with probability of the order of the B concentration, $L^{-d/2}$. Therefore the probability that an A particle reacts with any B particle before exiting this domain is of order $L^{(4-d)/2}$. Since this probability vanishes as $L \to \infty$ when $d > 4$, a domain is transparent to particles of the opposite species and the system therefore remains spatially homogeneous.

A naive way to simulate a diffusive reaction is to pick a reactant at random, allow it to diffuse a short distance (or hop to a random nearest-neighbor site), and perform a reaction if the particle that moved is within the reaction radius of another reactant. In an irreversible reaction where the density continuously decreases, such an approach becomes hopelessly inefficient. As the reactants become widely separated, most hopping events do not lead to a change in the reactant density. Thus a naive simulation spends most of the time performing diffusion steps and almost none of the time performing essential reaction steps.

An *event-driven* simulation mitigates this shortcoming, by ensuring that each update step corresponds to a reaction event. Let's specialize to the case of annihilation ($A + A \rightarrow \emptyset$) on a one-dimensional ring for simplicity; however, with a little more work the approach we now describe can be generalized to any spatial dimension. For a given spatial distribution of particles, we know, from Section 2.6 , the distribution of first-passage times for any nearest-neighbor pair to collide. Thus we may immediately advance the time by this first collision time and remove the corresponding reacting pair. All other particles propagate to their new positions according to the Gaussian probability distribution of free diffusion over this collision time. In this way, a single update step corresponds to a progressively increasing physical time increment as the system becomes more dilute.

An important moral is that one should never *ever* simulate the motion of individual random walks for any application. Because the first-passage properties of random walks are known, there is always a way to avoid constructing the paths of diffusing particles by mindless step-by-step hopping. Instead, the time update should be defined by the important physical events of the system.

13.4 Single-species reactions in one dimension

Diffusion-limited reactions in one dimension are special because they are often exactly soluble. In this section, we present exact solutions to two classic single-species reactions: (i) irreversible coalescence $A + A \rightarrow A$, and (ii) annihilation $A + A \rightarrow \emptyset$, both irreversible and in the steady state.

Irreversible coalescence $A + A \rightarrow A$

In coalescence, particles hop between nearest-neighbor sites with a rate that we set to $1/2$. Whenever a particle hops to an occupied site, the two particles coalesce to form a single particle (Fig. 13.4).

It is convenient to analyze this reaction by the *empty interval method* (see Chapter 7) that focuses on the voids between neighboring particles. The length of a void may either grow or shrink by 1 only because of the hopping of a particle at the end of a void. This fact allows us to write a soluble equation for the void length distribution. Let V_n be the density of voids of length n. As illustrated in Fig. 13.5, the length of a void undergoes a one-dimensional
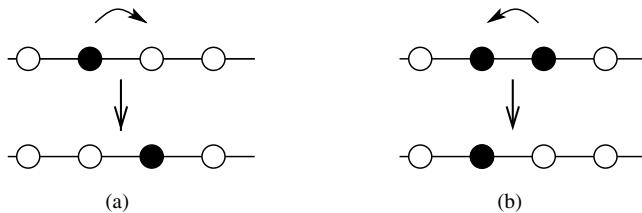
Fig. 13.4.   Coalescence in one dimension: (a) hopping to an empty site with rate 1/2, and (b) coalescence with rate 1.
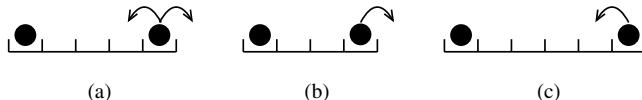
Fig. 13.5.   Changes in voids of length $n = 3$ due to hopping. Shown are the hopping events that lead to a loss (a) and the two types of gain processes (b) and (c).

random walk and the void length density evolves as (for $n > 0$)

$$\frac{dV_n}{dt} = V_{n+1} - 2V_n + V_{n-1}. \tag{13.29}$$

This equation can be extended to $n = 0$ by noting that the equation for the density of zero-length voids, $dV_0/dt = -2V_0 + V_1$, can be put into the same form as (13.29) by imposing the boundary condition $V_{-1} \equiv 0$.

As discussed in Section 2.1, the elemental solution to Eq. (13.29) has the form $I_n(2t)\, e^{-2t}$, with I_n the modified Bessel function of the first kind (see Eq. (2.15)). We now need the right combination of elemental solutions to satisfy the initial condition. For the canonical initial state in which every site is occupied, the initial condition for the voids is $V_n(0) = \delta_{n,0}$. To satisfy the boundary condition $V_{-1}(t) = 0$, we use the image method by initially placing a negative image of strength -1 at $n = -2$. By this construction, the boundary condition $V_{-1}(t) = 0$ is automatically satisfied. Superposing the initial and image contributions, the void density is

$$V_n(t) = \left[I_n(2t) - I_{n+2}(2t) \right] e^{-2t} = \frac{n+1}{t}\, I_{n+1}(2t)\, e^{-2t}\,, \tag{13.30}$$

where we use the identity $I_{n-1}(x) - I_{n+1}(x) = (2n/x) I_n(x)$ to obtain the final result.

Since particles and voids are in one-to-one correspondence, the particle density equals the density of voids of any length, $c(t) = \sum_{k \geq 0} V_k(t)$. Thus

$$c = \left[I_0(2t) + I_1(2t) \right] e^{-2t} \simeq \frac{1}{\sqrt{\pi t}}, \qquad t \to \infty. \tag{13.31}$$

The remarkable feature of this solution is that an interacting many-body system is reduced to a one-dimensional random walk with the right perspective. In the present case, the key

insight is that the void length corresponds to the position of the random walk. Essentially all of the exact solutions for one-dimensional reactions are variations on this theme.

From this analogy, we can immediately infer several basic properties of the void length distribution: (i) the average void length grows as $t^{1/2}$, (ii) the probability to have a long void of length $x \gg t^{1/2}$ decays as e^{-ax^2}, and (iii) the probability of a short void of length $x \ll t^{1/2}$ decays linearly with x. The latter feature is a manifestation of the effective reaction-induced repulsion between particles; if there was no interaction, the probability to have a small gap of length x would vary as $e^{-c(t)x}$.

Irreversible annihilation in one dimension

In irreversible annihilation, both particles disappear whenever a particle hops to an occupied site. If each particle hops with rate $1/2$, then annihilation occurs with rate 1, because either of the two adjacent particles may hop and cause a reaction. Irreversible annihilation is soluble because it is equivalent to the zero-temperature Ising–Glauber model (Section 8.5). Specifically, particle hopping is equivalent to an energy-conserving spin-flip event, while annihilation is equivalent to an energy-lowering spin flip (Fig. 13.6). Formally, the occupation number n_i at the ith site ($n_i = 1$ if the site is occupied and $n_i = 0$ otherwise) is obtained from the corresponding spin configuration on the dual lattice by $n_i = (1 - s_i s_{i+1})/2$. We may therefore simply borrow results from the Glauber solution to solve the annihilation reaction.

For an initially full lattice, $c(0) = 1$, corresponding to the antiferromagnetic initial state in the equivalent spin system, the time dependence of the concentration is (see Eq. (8.58)),

$$c(t) = I_0(2t)\, e^{-2t} \longrightarrow \frac{1}{\sqrt{4\pi t}} \qquad \text{as} \quad t \to \infty. \tag{13.32}$$

The asymptotic density is one-half that in coalescence (Eq. (13.31)), as two particles disappear in each annihilation event, while only one particle disappears in each coalescence. We can obtain the long-time behavior by the following physical argument: A random walk visits a region of size $\sqrt{Dt}$ after a time t. Because of annihilation, the number of particles that remain in this region must be of the order of 1; if there were more particles, they would have already annihilated. Thus the typical spacing between particles is of the order of $\sqrt{Dt}$,
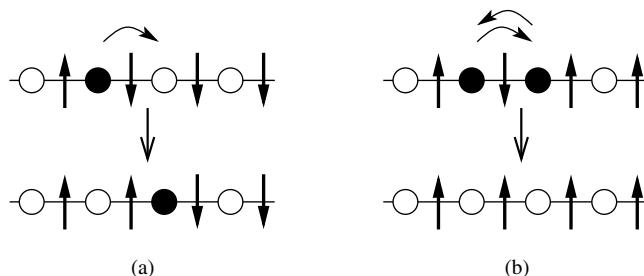
(a)   (b)

Single-species annihilation in one dimension: (a) hopping to an empty site with rate 1/2 and (b) annihilation with rate 1. Also shown is an equivalent representation in terms of Ising spins, where a pair of oppositely oriented spins is equivalent to an intervening domain-wall particle.

and the concentration is the inverse of this spacing. Notice that in both this rough argument and the exact result, the asymptotic density is independent of the initial density.

The main qualitative features of the void length distribution are known because of the correspondence with the domain length distribution in the Ising–Glauber model (Section 8.5, page 255). These features include: (i) the typical void length grows as $t^{1/2}$ (corresponding to the particle density decaying as $t^{-1/2}$); (ii) the density of small voids of length $x \ll t^{1/2}$ varies linearly with x; and (iii) the density of large voids of length $x \gg t^{1/2}$ decays exponentially with x. The exact solution for this distribution remains open.
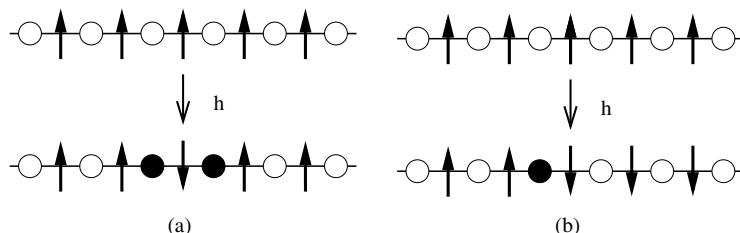
Steady-state annihilation

When particles are injected into the system at a constant rate, a steady state arises whose properties can again be exactly solved. A subtle feature is that the precise nature of the input determines whether the long-time state represents true thermodynamic equilibrium or a steady state. Let's first consider the injection of neighboring particle *pairs* at rate h, while the reaction $A + A \to \emptyset$ between neighboring particle pairs occurs at rate 1. Why consider this seemingly arcane example? From Fig. 13.7(a), pair input corresponds to an energy-raising single spin-flip event in the equivalent Ising–Glauber model. Consequently, annihilation with pair input corresponds to a non-zero temperature in the Ising–Glauber model and we can exploit known results from the latter to infer the properties of the reaction with input.

Because of the correspondence between pair input and an equilibrium spin system, detailed balance is satisfied. We may apply this condition to derive the equilibrium density. Since $A + A \to \emptyset\emptyset$ at rate 1 and $\emptyset\emptyset \to A + A$ at rate h, the detailed balance condition is

$$1 \times P(\ldots, 1, 1, \ldots) = h \times P(\ldots, 0, 0, \ldots),$$

where $P(\ldots, n_{i-1}, n_i, n_{i+1}, \ldots)$ denotes the probability of the occupancy configuration $\{n_i\}$. In thermal equilibrium, each pair of spins in the Ising model is independent and the distribution of domain walls factorizes into a product over individual domain-wall probabilities. Thus the above detailed balance condition becomes $c^2 = h(1-c)^2$, where c, the domain-wall

(a)   (b)
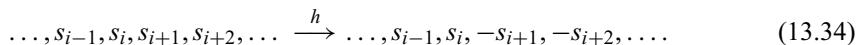
Fig. 13.7. Equivalence between the input of: (a) a pair of neighboring particles at rate h and an energy-raising event in the Ising–Glauber model, and (b) an input of a single particle at rate h and a non-local, energy-raising event in the Ising–Glauber model in which all spins to the right of the particle flip.

density, coincides with the particle density. The steady-state density is therefore

$$c = \frac{\sqrt{h}}{1 + \sqrt{h}} \, . \tag{13.33}$$

The above result is obvious *a posteriori* because neighboring sites are uncorrelated. For such pairs, the rate of particle gain equals the probability of finding two vacant sites times the creation rate; similarly, the particle loss rate equals the probability of finding two occupied sites times the annihilation rate. These two processes lead to the Langmuir-like evolution equation $dc/dt = -2c^2 + 2h(1 - c)^2$, whose equilibrium solution is (13.33).

The steady state has a markedly different nature when single particles are injected at rate h (Fig. 13.7(b)). For single-particle input, detailed balance cannot be satisfied as there are no processes that reverse the effect of annihilation and the input.[6] Thus the system reaches a steady state with a fundamentally non-equilibrium character. Nevertheless, we can determine the properties of this steady state by the Glauber formalism. When a single particle is created at site i, the occupation n_i changes from 0 to 1. In the equivalent Ising–Glauber model, this single-particle creation corresponds to flipping all spins to the right of the ith bond (Fig. 13.7(b)), that is,

$$\ldots, s_{i-1}, s_i, s_{i+1}, s_{i+2}, \ldots \xrightarrow{\;h\;} \ldots, s_{i-1}, s_i, -s_{i+1}, -s_{i+2}, \ldots . \tag{13.34}$$

Whenever a particle is added at the ith site, with $1 \leq i \leq k$, the product $g_k \equiv s_0 s_k$ changes sign. Therefore $g_k(t + \Delta t) = -g_k(t)$ with probability $hk\Delta t$, and $g_k(t + \Delta t) = g_k(t)$ with probability $1 - (hk)\Delta t$. Thus the rate of change in the correlation function $G_k = \langle g_k \rangle$ by the process in Eq. (13.34) equals $-2hkG_k$. We now add this term to the master equation for the correlation function of the Ising–Glauber model in Eq. (8.54) to give

$$\frac{dG_k}{dt} = -2(1 + kh)G_k + G_{k-1} + G_{k+1} \tag{13.35}$$

for $k \geq 1$. The boundary condition for this equation is $G_0 = 1$. In the steady state, Eq. (13.35) closely resembles the recursion relation for the Bessel function of the first kind[7]

$$\frac{2\nu}{x}J_\nu(x) = J_{\nu-1}(x) + J_{\nu+1}(x). \tag{13.36}$$

We now match this recursion with (13.35) by setting $2\nu/x = 2(1 + kh)$. This defines a one-parameter family of relations that connect (k, h) with (ν, x). A simple choice is $x = 1/h$ and $\nu = k + 1/h$, which then gives the pair correlation function for $k \geq 0$:

$$G_k(h) = \frac{J_{k+1/h}(1/h)}{J_{1/h}(1/h)} \, . \tag{13.37}$$

---

[6] If a particle is added to an already occupied site, then annihilation is defined to immediately occur so that the outcome is an empty site. The influence of this $A \to \emptyset$ process is negligible, however, in the interesting limit of $h \to 0$.

[7] Many of the calculational steps that follow are almost identical to those given in Section 7.5 for chaperone-assisted polymer translocation.

In the small-input limit, we use the asymptotic behavior of the Bessel function in terms of the Airy function, $J_\nu(\nu + x\nu^{1/3}) \simeq (2/\nu)^{1/3} \mathrm{Ai}(-2^{1/3}x)$ (Eq. (7.73)), to express the particle density as

$$c = \frac{1}{2}(1 - G_1) \simeq \frac{1}{2}\left[1 - \frac{\mathrm{Ai}((2h)^{1/3})}{\mathrm{Ai}(0)}\right], \qquad (13.38)$$

and expanding the Airy function to first order for small h gives

$$c \simeq C h^{1/3}, \qquad C = -2^{-2/3}\frac{\mathrm{Ai}'(0)}{\mathrm{Ai}(0)} = \left(\frac{3}{4}\right)^{1/3}\frac{\Gamma(2/3)}{\Gamma(1/3)} \doteq 0.459248. \qquad (13.39)$$

The steady-state densities given in Eqs (13.39) and (13.33) are counterintuitive at first sight. For a small rate h, pair input gives a density proportional to $h^{1/2}$, while single-particle input, with only half as many particles injected per unit time, gives a larger density proportional to $h^{1/3}$. However, the effective rate in pair input renormalizes to zero because neighboring particle pairs are certain to eventually annihilate.

Example 13.6. *Large-distance decay of correlations*. From the exact solution (13.37) for the pair correlation function, we only need to consider the $k \to \infty$ limit. It is easier, however, to deal with the governing equation in the asymptotic regime, rather than take the asymptotic limit of the exact solution. In the $k \to \infty$ limit, we can replace $G_{k+1} - 2G_k + G_{k-1}$ by d^2G/dk^2, so that in the steady state Eq. (13.35) becomes

$$\frac{d^2G}{\partial k^2} = 2hk\, G. \qquad (13.40)$$

The asymptotic behavior of solutions to this differential equation[8] is easy to establish using the WKB method (see the highlight in Section 7.5, page 229). The result is

$$G_k \sim \exp\left[-\sqrt{8h/9}\; k^{3/2}\right].$$

This decay is unexpectedly quick – it is faster than the exponential decay of Eq. (8.55) that characterizes the equilibrium Ising model.

As a counterpoint to exact analysis, let's try to understand the steady state and the approach to this state by scaling arguments. In the small-input limit, reaction is initially the dominant mechanism. Eventually the density decays to a level where the reaction and input rates are comparable. We define the time at which the density, which initially decays as $t^{-1/2}$, crosses over to a constant level as $t^* \sim h^{-\beta}$. These two limiting behaviors are encapsulated by the scaling ansatz

$$c(h,t) \sim h^\alpha\, \Phi(h^\beta\, t) \qquad \text{with} \quad \alpha = \begin{cases} 1/3, & \text{pair input,} \\ 1/2, & \text{single-particle input.} \end{cases} \qquad (13.41)$$

---

[8] Align (13.40) is known as the Airy equation.

To match to the temporal decay of $c \sim t^{-1/2}$ for $t \ll t^*$, the scaling function must have the power-law tail $\Phi(z) \sim z^{-1/2}$ as $z \to 0$. This dependence implies the exponent relation $\beta = 2\alpha$, so as to eliminate the dependence on the input rate for sufficiently short times. Thus we conclude that

$$\beta = \begin{cases} 1, & \text{pair input,} \\ 2/3, & \text{single-particle input.} \end{cases} \tag{13.42}$$

The crossover to the steady state occurs at $t^* \sim h^{-1}$ for pair input, which is substantially later than $t^* \sim h^{-2/3}$ for single-particle input.

One-dimensional aggregation

Aggregation in one dimension can be solved analytically by a generalization of the empty interval method. This approach is applicable only in the idealized situation where clusters hop with mass-independent rate. We additionally assume that each cluster occupies a single site. When a cluster hops onto a site that is occupied by another cluster, the clusters immediately join to form a larger cluster: $A_i + A_j \to A_{i+j}$.

Irreversible aggregation

Let $c_k(t)$ be the density of clusters of mass k at time t. We consider the initial condition in which every site is initially occupied by a monomer, $c_k(0) = \delta_{k,1}$, so that the properties of the system are translationally invariant. To solve for c_k, it is convenient to introduce the probability Q_n^k that the total mass contained in n consecutive sites equals k. By construction $\sum_{k \geq 0} Q_n^k = 1$, and Q_n^0 is the probability that there is no mass in an interval of length n. Thus Q_n^0 coincides with the empty interval probability that was introduced in Chapter 7 on adsorption phenomena: $E_n = Q_n^0$. The cluster mass distribution simply equals the probability that a mass k is contained in an interval of length 1; that is, $c_k = Q_1^k$.

The feature that renders aggregation solvable is that the interval occupancy probabilities Q_n^k evolve according to the same discrete diffusion equation that governs the void density in single-species coalescence. The derivation, however, is a bit more delicate than that for the void density because we need to account for both the interval length and the contained mass. To write the master equation for Q_n^k, we will also need the conditional probability $\widehat{Q}_n^k$ that n consecutive sites contain a total mass k *and* the next site is empty. In terms of these conditional probabilities, we now enumerate the ways that Q_n^k can change due to hopping events at the right boundary (Fig. 13.8); a similar set of contributions arises at the left boundary. Intervals of size n and mass k are gained ($+$) and lost ($-$) with the following rates:

+ The mass in an n-interval is smaller than k and a cluster hops into this interval to make the final mass equal to k. Thus the mass contained in the interval of length $n+1$ must equal k. The rate of this event equals $[Q_{n+1}^k - \widehat{Q}_{n+1}^k]/2$. The difference of the Qs accounts for the probability that a mass k is contained in an interval of length $n+1$ with the rightmost
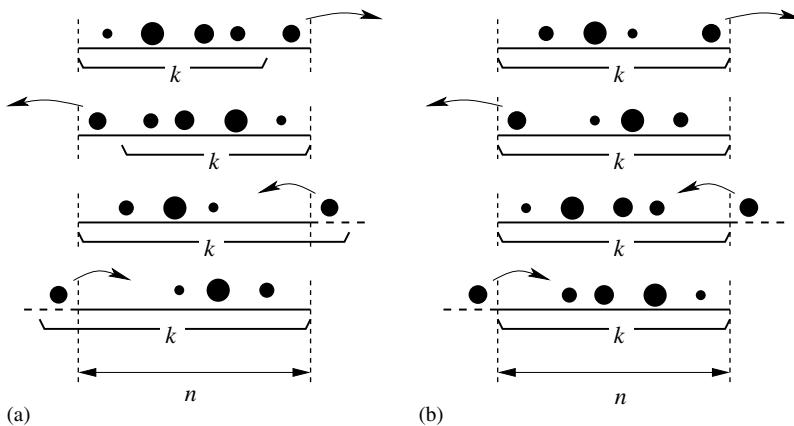
Fig. 13.8.   Configurations that contribute to the change in the set of states in which a total mass k is contained in an interval of length n: (a) the four processes that lead to a gain in Q_n^k, and (b) the four loss processes.

site of this interval being occupied. The factor 1/2 accounts for this last cluster hopping to the left to create an interval of length n that contains mass k.

+ The interval mass is larger than k and a cluster at the end of the interval hops out so that the final mass equals k. The rate for this event is $[Q_{n-1}^k - \widehat{Q}_{n-1}^k]/2$.

− The interval mass equals k and a cluster hops into it with rate $-[Q_n^k - \widehat{Q}_{n+1}^k]/2$.

− The interval mass equals k and a cluster hops out of it with rate $-[Q_n^k - \widehat{Q}_{n-1}^k]/2$.

Adding these transition rates, the conditional probabilities miraculously cancel! By including an identical set of contributions due to the left boundary, the evolution of the empty interval density is again described by the discrete diffusion equation

$$\frac{dQ_n^k}{dt} = Q_{n+1}^k - 2Q_n^k + Q_{n-1}^k \tag{13.43}$$

for all $k \geq 0$. The boundary condition is $Q_0^k(t) = 0$ for $k > 0$ (indeed, $E_0 = Q_0^0 = 1$) and the initial condition that each lattice site is initially occupied by a monomer leads to $Q_n^k(0) = \delta_{n,k}$. A nice feature of Eq. (13.43) is that different masses are uncoupled, so that the solution for each k can again be obtained by the image method (see Eq. (13.30)) to give $Q_n^k(t) = [I_{n-k}(2t) + I_{n+k}(2t)]e^{-2t}$, where I_m is again the modified Bessel function of order m. From this result, the cluster mass density $c_k = Q_1^k$ is

$$c_k(t) = [I_{k-1}(2t) - I_{k+1}(2t)]e^{-2t} \simeq \frac{k}{\sqrt{4\pi t^3}} e^{-k^2/4t}, \quad t \to \infty. \tag{13.44}$$

As an amusing postscript, this cluster mass density contains the solutions to single-species coalescence and single-species annihilation. The density of clusters of any mass in aggregation is identical to the cluster density c_{coa} in coalescence; that is, $c_{\mathrm{coa}} = c_1 + c_2 + c_3 + \cdots$.

Similarly, if the system initially contains only monomers, the cluster density in single-species annihilation, c_{ann}, coincides with the density of odd-mass clusters in aggregation; hence $c_{ann} = c_1 + c_3 + c_5 + \cdots$.

The requirement that the diffusion coefficients of clusters are mass-independent, $D_k \equiv D$, is crucial for the applicability of the mathematical framework given above; it is unknown how to deal with mass-dependent diffusion coefficients. The best we can do is to use scaling arguments to estimate the temporal evolution of the cluster density.

Example 13.7. *Aggregation with mass-dependent diffusion coefficients.* In the simplest case of mass-independent diffusion, a diffusing particle spans an interval of length of the order of $\sqrt{Dt}$ during the time interval $(0, t)$. Hence each such interval contains on average one cluster; therefore, the average cluster mass is

$$m \sim \sqrt{Dt} \,. \tag{13.45}$$

Consider the situation when diffusion coefficients scale algebraically with mass, say $D_k \sim k^{-\nu}$. The scaling hypothesis suggests characterizing the cluster population by a typical mass. Thus we substitute $D = D_m \sim m^{-\nu}$ into (13.45) to give $m \sim t^{1/(2+\nu)}$. Accordingly, the cluster density decays as

$$c \sim m^{-1} \sim t^{-1/(2+\nu)} \,. \tag{13.46}$$

We further expect that the mass distribution approaches the scaling form, $c_k(t) = c^2 F(ck)$. However, our heuristic argument that led to the density decay (13.46) is insufficient to make conclusions about the behavior of the scaled mass distribution function F.

Aggregation with input

What happens when we add monomers to the system spatially uniformly and at a constant rate? We previously found that, for the constant reaction kernel in the mean-field limit, the steady-state mass distribution is $c_k(t \to \infty) \sim k^{-3/2}$ over a mass range $1 \ll k \ll t^2$ (Section 5.5). Let's determine the corresponding behavior in one dimension.

When monomers are added to each site at a constant rate h, the master equations for the empty interval probabilities are still soluble. The effect of input on an empty interval probability is simple: if an n-interval contains mass $k - 1$ and an input event occurs, there is a gain in the interval occupancy probability Q_n^k. The rate at which mass is added to an n-interval is $h\,n$ because input may occur at any of the sites in the interval. Similarly, if the interval contains mass k, input causes the loss of Q_n^k at rate $h\,n$. Thus the equation (13.43) for Q_n^k generalizes to

$$\frac{dQ_n^k}{dt} = -2Q_n^k + Q_{n-1}^k + Q_{n+1}^k + h\,n \left[Q_n^{k-1} - Q_n^k \right]. \tag{13.47}$$

This equation holds for all $k \geq 0$ with the boundary condition $Q_n^{-1}(t) = 0$.

Random river networks

Aggregation with input is equivalent to the classic Scheidegger random river network model (Fig. 13.9). In this model, the downstream position along a river is equivalent to the time and its lateral meandering is equivalent to a one-dimensional random walk. The merging of two rivers corresponds to aggregation. The area of a river basin (dashed lines in Fig. 13.9) corresponds to the flow at a river mouth. This flow is equivalent to the mass of the aggregate in the space-time diagram, when there is also a constant input of mass at each site of the one-dimensional lattice.
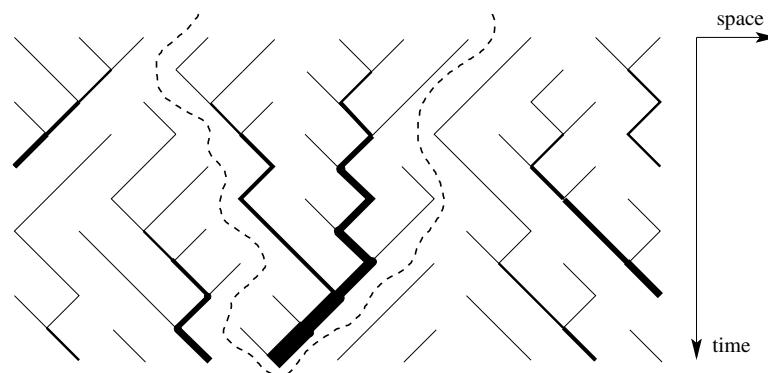
We can exploit this picture to give a heuristic argument for the mass distribution of the aggregates. The area of a given basin is proportional to its height h times its width. Since the trajectories of the boundaries of a river basin perform random walks, the width should scale as $h^{1/2}$ and the area scales as $h^{3/2}$. This area corresponds to the mass k of an aggregate. On the other hand, h can be viewed as the time at which two random walks first meet, a probability that asymptotically decays as $p(h) \sim h^{-3/2}$. By transforming from h to k, the mass distribution is given by

$$c_k = p(h)\frac{dh}{dk} \sim h^{-3/2}\frac{dh}{dk} \sim k^{-1}k^{-1/3} \sim k^{-4/3}\,, \tag{13.48}$$

which agrees with the asymptotic result (13.52).

These equations couple the interval probabilities for different masses, as well as different lengths; thus their solution is more complex than the mass-decoupled equations in the absence of input. However, the situation is simpler in the steady state. In this case, we use the generating function, $Q_n(z) = \sum_{k\geq 0} Q_n^k z^k$, to convert Eq. (13.47) to

$$Q_{n-1}(z) + Q_{n+1}(z) = [2 + hn(1-z)]\,Q_n(z)\,. \tag{13.49}$$

Fig. 13.9. Space-time representation of a discrete-time version of aggregation with input. The 45° line segments represent the trajectories of aggregates that perform synchronous random walks in one dimension. In the river representation, a wider river is formed when two tributaries meet. Input is accounted for by defining the mass of an aggregate as the area of the corresponding river basin.

This recursion formula is the same as that for the Bessel function (see Eq. (13.36)) when the index is properly matched. Thus following the same line of reasoning as that leading to Eq. (13.37), the solution for Q_n is

$$Q_n(z) = \frac{J_{n+1/g}(1/g)}{J_{1/g}(1/g)}, \tag{13.50}$$

with $g \equiv g(z, h) = h(1 - z)/2$. We are interested in the large-k behavior of the mass distribution, corresponding to the $z \to 1$ limit of the generating function. In this situation $g \ll 1$ and we can use the asymptotic formula (7.73) for the Bessel function to obtain

$$Q_1(z) \simeq \frac{\mathrm{Ai}((2g)^{1/3})}{\mathrm{Ai}(0)} \simeq 1 + \frac{\mathrm{Ai}'(0)}{\mathrm{Ai}(0)} h^{1/3} (1 - z)^{1/3}. \tag{13.51}$$

The singular $(1 - z)^{1/3}$ behavior of the generating function implies (see Eq. (5.26) in the highlight on page 142) a $k^{-4/3}$ tail of the mass distribution. More precisely, for $k \gg 1$ we obtain

$$c_k \simeq C h^{1/3} k^{-4/3}, \qquad C = \frac{\mathrm{Ai}'(0)}{\Gamma(-1/3)\,\mathrm{Ai}(0)} = \frac{1}{3^{2/3}\,\Gamma(1/3)} \doteq 0.179455. \tag{13.52}$$

The entire mass distribution is encapsulated in Eq. (13.50) at $n = 1$, or equivalently

$$1 + \sum_{k \geq 1} c_k(z^k - 1) = \frac{J_{1+1/g}(1/g)}{J_{1/g}(1/g)}, \qquad g = \frac{h(1 - z)}{2}. \tag{13.53}$$

Explicit results are easy to extract in the small-input limit. Indeed, when $h \to 0$, the right-hand sides of Eqs (13.53) and (13.51) coincide. Therefore

$$\sum_{k \geq 1} c_k(z^k - 1) = \frac{\mathrm{Ai}'(0)}{\mathrm{Ai}(0)} h^{1/3} (1 - z)^{1/3} \tag{13.54}$$

when $h \ll 1$. In contrast to Eq. (13.51), which is valid in the $z \to 1$ limit, Eq. (13.54) applies for all z. Using (13.54) in conjunction with the binomial identity,

$$(1 - z)^a = \sum_{k \geq 0} \frac{\Gamma(k - a)}{\Gamma(-a)\,\Gamma(k + 1)} z^k,$$

we determine the entire mass distribution

$$c_k = C h^{1/3} \frac{\Gamma(k - 1/3)}{\Gamma(k + 1)}, \tag{13.55}$$

with amplitude C the same as in (13.52).

13.5 Reactions in spatial gradients

Logistic growth and the FKPP equation

Let's start by studying the role of diffusion in the idealized logistic growth model (Section 12.1), in which the competition between birth and death in a homogeneous population leads to a steady state. When diffusion is incorporated, we now address the following basic question: How does a small and spatially localized population spread throughout the system? This "invasion" of a stable unit-density population into an unpopulated region is mathematically described by the Fisher–Kolmogorov–Petrovsky–Piscounov (FKPP) equation

$$\frac{\partial A}{\partial t} = D\nabla^2 A + kA(1 - A). \tag{13.56a}$$

The reaction terms describe the evolution of a population by logistic growth in which birth ($A \to A + A$) is counterbalanced by self-regulation ($A + A \to A$). The first term on the right-hand side accounts for the diffusion of reactants. As discussed in Section 12.1, in the absence of the diffusion term (which corresponds physically to an infinite diffusion constant), the resulting rate equation, $\dot{A} = A(1 - A)$, has a stable fixed point when $A = 1$ and an unstable fixed point for $A = 0$. Thus starting with any non-zero density, a final density of 1 is reached.

To describe the invasion of an empty region by the stable state, consider the simple setting in which the density $A = 1$ in the half-space $x < 0$, and $A = 0$ for $x > 0$ (with no dependence on transverse coordinates). By diffusion alone, an initially sharp interface between these two regions becomes smoother, with an error function density profile. However, the combination of diffusion and reaction causes the stable high-density phase to propagate at constant velocity into the low-density phase (Fig. 13.10). To understand this invasion dynamics, it is first useful to rescale the time by $t \to kt$ and the length by $x \to x\sqrt{k/D}$ to non-dimensionalize the FKPP equation to

$$\frac{\partial A}{\partial t} = \nabla^2 A + A(1 - A). \tag{13.56b}$$
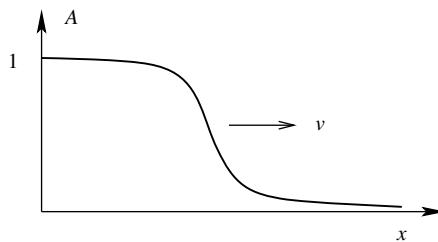
Fig. 13.10. Invasion of an unstable region with density $A = 0$ by a stable population with $A = 1$.

Let us now assume that the stable state propagates as a wave of stationary shape into the unstable phase. We further assume that the waveform has an asymptotic exponential decay so that

$$A(x, t) \sim e^{-\lambda(x - vt)} .$$

Substituting this form into the dimensionless FKPP equation (13.56b) and dropping the nonlinear term that is negligible at the leading edge of the wave, we obtain the dispersion relation

$$v(\lambda) = \lambda + \lambda^{-1} . \tag{13.57}$$

Thus the crudest assumption about the nature of the propagation gives a non-trivial relation between the propagation speed and the shape of the wave. However, Eq. (13.57) is incomplete because we can only surmise that the velocity depends on λ and satisfies the bound $v \geq v_{\min} = 2$, where $v_{\min}$ is the minimum possible velocity from the dispersion relation in (13.57). Which velocity is actually selected is undetermined. Nature is wise, however; for reasonable initial conditions, $v = v_{\min}$ is selected. Here "reasonable" means an initial profile with a compact support, that is, vanishing beyond a certain threshold, or with a sufficient rapidly decaying tail. Specifically, for initial profiles with an exponential decay, $A(x, 0) \sim e^{-\lambda x}$ for large x, the position x_f of the wavefront advances with the minimum velocity only when λ is sufficiently large. The precise statement is

$$x_f = \begin{cases} 2t - \frac{3}{2} \ln t + \mathcal{O}(1), & \lambda > 1, \\ 2t - \frac{1}{2} \ln t + \mathcal{O}(1), & \lambda = 1, \\ v(\lambda) t + \mathcal{O}(1), & \lambda < 1, \end{cases}$$

with $v(\lambda) = \lambda + \lambda^{-1} > 2$ when $\lambda < 1$. Moreover, if the initial density profile decays as $e^{-\lambda_0 x}$ for large x with $\lambda_0 > 1$, then the leading edge of the wave asymptotically evolves to an exponential with decay parameter $\lambda = 1$ and velocity $v = v_{\min} = 2$. On the other hand, if $\lambda_0 < 1$, then the leading edge of the wave preserves this slower decay, $e^{-\lambda_0(x - vt)}$, and the velocity is $v = \lambda_0 + \lambda_0^{-1} > v_{\min}$.

Aggregation with spatially localized input

Another setting where spatial gradients affect reaction kinetics is aggregation in the presence of a *spatially localized* input – for example, constant injection of monomers at a single point. For this system, the governing equations for the cluster distribution are

$$\frac{\partial c_k}{\partial t} = D\nabla^2 c_k + K \sum_{i+j=k} c_i c_j - 2Kc_k N + J\delta_{k,1}\delta(\mathbf{r}) . \tag{13.58}$$

Here $c_k(\mathbf{r}, t)$ is the density of clusters of mass k at $(\mathbf{r}, t)$. The first term on the right-hand side of Eq. (13.58) describes mixing due to diffusion, the next two terms account for constant-kernel aggregation, and the last term is for monomer input at the origin at rate J.

Let's start by studying the evolution of the spatially dependent mass density $M(\mathbf{r}, t) = \sum_k kc_k(\mathbf{r}, t)$. Since aggregation is a mass-conserving process, the reaction does not affect

the mass distribution. The mass density obeys the diffusion equation augmented by the source term:

$$\frac{\partial M}{\partial t} = D\nabla^2 M + J\delta(\mathbf{r}).$$ (13.59)

Suppose that the source is turned on at $t = 0$ in an initially empty system. Then the mass distribution is simply the superposition of Gaussian propagators due to all sources from $t = 0$ until the present:

$$M(\mathbf{r}, t) = J \int_0^t \frac{dt'}{(4\pi Dt')^{d/2}} \, e^{-r^2/4Dt'}.$$ (13.60)

The solution to Eq. (13.60) in the limits $t \to \infty$ and $r \ll \sqrt{Dt}$ leads to the mass density (problem 13.5):

$$M(\mathbf{r}) = \frac{J}{4\pi D} \times \begin{cases} 4\sqrt{\pi Dt}, & d = 1, \\ \ln(4Dt/r^2) - \gamma_E + \mathcal{O}(r^2/4Dt), & d = 2, \\ 1/r, & d = 3, \end{cases}$$ (13.61)

where $\gamma_E \doteq 0.577215$ is Euler's constant. Notice that the mass density reaches a stationary limit only when the spatial dimension $d > 2$; in this situation, the mass density is just given by the solution to the (time-independent) Laplace equation.[9]

Next consider the density of clusters, $N(\mathbf{r}, t) = \sum_{k\geq 1} c_k(\mathbf{r}, t)$. Since we use the rate equations (13.58) which are valid above the critical dimension ($d_c = 2$ in the case of constant-kernel aggregation), we tacitly assume that the underlying space is three-dimensional. However, the effective dimension can be smaller by considering the following geometries:

- uniform injection on the plane $x = 0$ in $d = 3$, an effective one-dimensional system;
- uniform injection along the line $x = y = 0$ in $d = 3$, an effective two-dimensional system;
- injection at the origin $x = y = z = 0$ for $d = 3$.

We shall also analyze what happens when $d > 3$. As we've learned previously, such a generalization helps to put results in the proper perspective and enhances our understanding of the physically relevant three-dimensional situation.

The governing equation for the cluster density is obtained by summing Eqs (13.58). In the $t \to \infty$ limit we obtain the reaction–diffusion equation

$$D\nabla^2 N - KN^2 + J\delta(\mathbf{r}) = 0.$$ (13.62)

Here we tacitly assume that the cluster density reaches a stationary limit. This stationarity is not self-evident; we have seen that the mass density is not stationary when $d \leq 2$. In

---

[9] This stationarity reflects a crucial property of diffusion: it is transient in $d > 2$ and therefore there is no accumulation of particles at the origin. Conversely, for $d \leq 2$, diffusion is recurrent and the density near the origin grows without bound because previously created particles return infinitely often.

principle, we should consider the time-dependent version of (13.62), solve this equation, and show that this solution reaches a stationary limit. While a procedure was feasible for the mass density, it is not feasible for the time-dependent version of (13.62). While we cannot describe $N(\mathbf{r}, t)$ analytically, we can try to understand its asymptotic behavior by assuming stationarity and checking consistency *a posteriori*.

Stationary solutions of Eq. (13.62) exhibit an interesting dependence on spatial dimension. When d exceeds an upper critical dimension d^c (which we will determine later from consistency), the reaction term is negligible far away from the source. Hence when $r \to \infty$, Eq. (13.62) reduces to the Laplace equation, with asymptotic solution

$$N \sim \frac{J}{D} \frac{1}{r^{d-2}} \qquad \text{when} \quad d > d^c .\qquad (13.63\text{a})$$

Conversely, when we cannot neglect the reaction term in (13.62), we use simple power counting to determine the r dependence of the density. Writing $\nabla^2 N \sim N/r^2 \sim N^2$, we infer that $N \sim r^{-2}$. Thus we substitute the ansatz $N = A/r^2$ into (13.62). A straightforward calculation gives an asymptotic solution

$$N \simeq (4 - d) \frac{2D}{K} \frac{1}{r^2} \qquad \text{when} \quad d < d^c .\qquad (13.63\text{b})$$

Because the r dependences for $d > d^c$ and $d < d^c$ coincide in four dimensions, we deduce that the upper critical dimension is $d^c = 4$.

At $d = d^c = 4$, it is natural to expect logarithmic corrections. Thus we seek a solution for (13.62) of the form $N = u/r^2$, with $u = u(\rho)$ where $\rho = \ln r$. We expect u to vanish when $\rho \to \infty$ since, according to (13.63b), the amplitude vanishes as $d \uparrow d^c = 4$. In four dimensions,

$$\nabla^2 N = \frac{d^2 N}{dr^2} + \frac{3}{r} \frac{dN}{dr} .$$

We now substitute $N = u(\rho)/r^2$ and use

$$\frac{du}{dr} = \frac{u'}{r} , \qquad\qquad \frac{d^2 u}{dr^2} = -\frac{2u'}{r^2} + \frac{u''}{r^2} ,$$

where the prime denotes differentiation with respect to ρ, to recast Eq. (13.62) into

$$-2u' + u'' = \frac{K}{D} u^2 .$$

Since $u \to 0$ as $\rho \to \infty$, we guess that $u'' \ll u'$ and keep only the u' term.[10] This yields $u' = -Ku^2/2D$ and therefore in four dimensions the cluster density decays as

$$N \simeq \frac{2D}{K} \frac{1}{r^2 \ln r} \qquad\qquad (13.63\text{c})$$

far away from the source.

---

[10] This guess is correct for functions that decay slower than exponentially (e.g. algebraically). To be consistent, we must verify the inequality $u'' \ll u'$ *a posteriori*. For the present case we get $u \sim \rho^{-1}$ and therefore $u'' \sim \rho^{-3}$ is indeed negligible compared to $u' \sim \rho^{-2}$.

The results (13.63a)–(13.63c) are asymptotic. We now show that in one dimension the stationary problem admits an exact solution.

Example 13.8. *Cluster density in one dimension.* Multiplying (13.62) by $2N'$ and integrating from a non-zero value of x to $x = \infty$ gives

$$D(N')^2 - \frac{2}{3} KN^3 = 0, \tag{13.64}$$

where the constant on the right-hand side is zero since $N \to 0$ as $|x| \to \infty$. Integrating (13.64) we find

$$N = \frac{6D}{K} \frac{1}{(|x| + x_0)^2}. \tag{13.65}$$

To determine the integration constant x_0 we integrate (13.62) over an infinitesimal region $(-\epsilon, \epsilon)$ around the origin and let $\epsilon \to 0$. This procedure gives $D\left[N'(+0) - N'(-0)\right] + J = 0$, and using (13.65) we obtain

$$x_0 = \left(24 \frac{D^2}{JK}\right)^{1/3}. \tag{13.66}$$

The amplitude $6D/K$ in (13.65) agrees with the asymptotic solution (13.63b).

The exact solution (13.65) crosses over to the asymptotic regime when $|x| \sim x_0$. In two and three dimensions we anticipate a similar crossover at $r \sim r_0$. How does r_0 depend on the parameters D, K, and J of the problem? Unfortunately, dimensional analysis is insufficient to deal with this question. For example, in three dimensions we could use $(D/J)^{1/2}$ and $(K/J)^{1/3}$ as independent units with the dimension of length. A trick is helpful here. Guided by the form of the solution in one dimension, let us seek $N(r)$ in the form

$$N(r) = \frac{D}{K} f(r). \tag{13.67}$$

Substituting this ansatz into Eq. (13.62) we obtain

$$\nabla^2 f - f^2 + \frac{KJ}{D^2} \delta(\mathbf{r}) = 0. \tag{13.68}$$

This equation shows that, and instead of three variables D, K, and J, the solution actually depends on the single combination KJ/D^2. This combination has the dimension of length raised to the power $d - 4$ and leads to the estimate

$$r_0 \sim \left(\frac{D^2}{KJ}\right)^{1/(4-d)}, \qquad d < 4. \tag{13.69}$$

Finally, let's determine the total number of clusters. In principle,

$$\mathcal{N}(t) = \int_0^\infty N(r, t)\, \Omega_d\, r^{d-1}\, dr. \tag{13.70}$$

In one dimension, we use (in the $t \to \infty$ limit) the exact stationary solution (13.65) and obtain

$$\mathcal{N}(t \to \infty) = \frac{6D}{K} \frac{1}{x_0} = \left(\frac{72\,DJ}{K^2}\right)^{1/3}. \tag{13.71}$$

Thus in one dimension the total number of clusters saturates to a finite value.

In two and three dimensions, we cannot merely insert the stationary cluster density into the right-hand side of Eq. (13.70) because the integral diverges in the $r \to \infty$ limit. However, since clusters perform random walks with a (mass-independent) diffusion coefficient D, the density of clusters quickly vanishes for $r \gg \sqrt{Dt}$. Thus to estimate the total number of clusters we can use the stationary results but cut off the integration at $\sqrt{Dt}$. Additionally, we need to cut off the lower limit at r_0. With these steps, we recast Eq. (13.70) into

$$\mathcal{N}(t) \sim \int_{r_0}^{\sqrt{Dt}} N(r)\, \Omega_d\, r^{d-1}\, dr. \tag{13.72}$$

Using this together with (13.63b) we find

$$\mathcal{N}(t) \sim \frac{D}{K} \times \begin{cases} \ln(KJt/D), & d = 2, \\ \sqrt{Dt}, & d = 3. \end{cases} \tag{13.73}$$
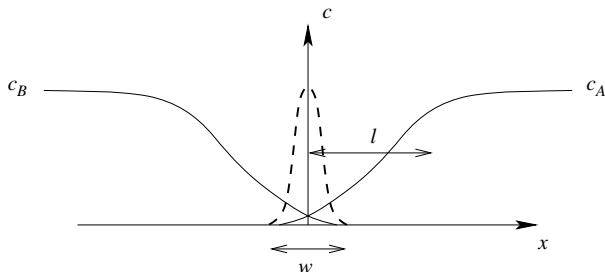
It is remarkable how slightly the source strength affects the cluster density and the total number of clusters in this source-driven aggregation process.

Inhomogeneous two-species annihilation

Let's return to the basic two-species annihilation reaction from Section 13.3, but instead of uniform initial densities let us consider spatially separated reactants. Because the opposite two species cannot coexist, it is more physical to study this reaction when the reactants are initially separated and are then brought into proximity. Here we discuss this reaction when two initially separated components react at a planar interface. The complementary situation of the steady response that arises when the two species are continuously fed into the system from localized sources is the subject of problems 13.9 and 13.10.

Consider a d-dimensional system that is divided by a $(d-1)$-dimensional hyperplane at $x = 0$, with As initially distributed at random in the region $x > 0$ and similarly for Bs in the region $x < 0$ (Fig. 13.11). The subsequent reaction kinetics is described by the reaction–diffusion equations

$$\begin{aligned} \frac{\partial c_A}{\partial t} &= D\, \frac{\partial^2 c_A}{\partial x^2} - Kc_Ac_B, \\ \frac{\partial c_B}{\partial t} &= D\, \frac{\partial^2 c_B}{\partial x^2} - Kc_Ac_B, \end{aligned} \tag{13.74}$$

Fig. 13.11. Sketch of the concentration profiles (solid curves) and the reaction rate $Kc_A c_B$ (dashed, and considerably magnified) in two-species annihilation with initially separated components. The depletion zone width is $\ell \sim \sqrt{Dt}$, while the reaction zone width is $w \sim t^{1/6}$.

where $c_A = c_A(x, t)$ and $c_B = c_B(x, t)$ denote the concentration of each species at position x at time t, and K is the reaction constant. This initial densities are

$$c_A(x, t = 0) = \begin{cases} c_0, & x > 0, \\ 0, & x < 0, \end{cases}$$

and $c_B(x, t = 0) = c_A(-x, t = 0)$. The basic features of this reaction are the width w of the reaction zone and the rate at which reactants are produced. As we now discuss, these properties can be determined by dimensional analysis alone.

For the geometry of Fig. 13.11, the domain of one species acts as a nearly fixed absorbing boundary condition for the opposite species when the reaction constant K is large. Thus the density profile of the As, for example, is close to that of independent diffusing particles in the presence of an absorbing boundary at $x = 0$, namely $c_A(x, t) = c_0 \, \mathrm{erf}(x/\sqrt{4Dt})$. From the asymptotics of the error function, the density varies as $c(x, t) \sim c_0 \, x/\sqrt{4Dt}$ for $0 < x < \sqrt{4Dt}$. Thus the size of the depletion zone is $\ell \sim \sqrt{Dt}$.

We may now estimate the width w of the reaction zone by balancing the diffusive flux into this zone, $J = 2|Dc'| = 2c_0\sqrt{D/t}$, with the rate at which particles are annihilated. Using the reaction–diffusion equation (13.74), the number of reactions per unit time equals the integral of $K \, c_A \, c_B$ over the extent of the reaction zone. We estimate this integral as the square of a typical concentration within the zone, namely, $c_A(x = w, t)$, times the zone width w. Using the small-x form $c_A \sim c_0 x/\sqrt{4Dt}$, the total reaction rate therefore is $Kc(w, t)^2 w \sim Kc_0^2 w^3/(Dt)$. Equating this rate to the input flux $c_0\sqrt{D/t}$ gives the width of the reaction zone:

$$w \sim (D^3 t/K^2 c_0^2)^{1/6} . \tag{13.75}$$

Correspondingly, the local production rate within the reaction zone, $Kc(w, t)^2$, vanishes as $t^{-2/3}$.

There are many interesting variations of this system. For example, when the reaction rate is sufficiently small, the initially separated species will first penetrate over a length scale that grows as $\sqrt{Dt}$. Reactions will subsequently occur homogeneously in this mixing

region, leading to an initial reaction rate that also grows as $\sqrt{Dt}$ before the asymptotic $t^{-2/3}$ decay sets in. The motion of the interface can also be non-monotonic in time when $D_A > D_B$ but $c_A(0) < c_B(0)$. At early times, the faster diffusion of the As dominates, and the interface moves to the left. At long times, however, the interface moves to the right because the majority species (B) overwhelms the A.

13.6 Notes

One of the salient features of diffusion-controlled reactions is that the spatial dimension plays a crucial role in their kinetics. This insight was developed through investigations of many disparate examples. Self-avoiding random walks are described in [11, 234]. The feature that $d_c = 1$ for three-body reactions is discussed in [235, 236]. The optimal fluctuation argument for the asymptotic decay of the density in the trapping reaction was developed by Balagurov and Vaks [237]; see also [238]. These works showed that departures from mean-field theory arise in all spatial dimensions, although the time needed to observe the true asymptotic behavior is astronomically large. The exact asymptotics for the density was found by Donsker and Varadhan [239]. The feature that the ball has the lowest eigenvalue of the Laplace–Dirichlet operator from all simply connected sets of fixed volume is known as the Faber–Krahn theorem that was proved in the 1920s for two dimensions. This theorem is described in [240]. Subsequently, this theorem was generalized to arbitrary dimensions and the proof can be found in [241].

The anomalously slow kinetics in two-species annihilation was recognized by Zeldovich [242] in 1949. The subject of diffusion-controlled reactions became quite popular in the late 1970s through the 1980s. A sampling of the important early work is [243, 244]. Some pioneering papers appeared in obscure journals in the Soviet literature; fortunately, some of this work is reviewed in [245]. Reviews of spatial organization effects in two-species annihilation are given in [246, 247]. The rigorous result for the decay of the density in two-species annihilation was derived by Bramson and Lebowitz [248]. The effort to develop efficient simulations of reactions has a long history. An early work to construct the first-passage method mentioned in the highlight at the end of Section 13.3 appears in [249]; this approach has been carried to a high art in [250]

The exact solutions of one-dimensional single-species reaction processes are discussed in the review articles that appear in [247]. The exact solution of one-dimensional annihilation in the steady state and the connection to the kinetic Ising model was given by Rácz [251]. For irreversible aggregation in one dimension, an exact solution for the cluster mass distribution was given by Spouge [252]. This approach was developed further by Thomson [253] and by Takayasu *et al.* [254] to treat aggregation with a steady input of monomers.

The phenomenon of the invasion of a stable population into an unstable region was treated by Kolmogorov *et al.* [255] and by Fisher [256]; a pedagogical discussion of this important example appears in Murray's book on mathematical biology [212]. The phenomenon of aggregation with a localized monomer source is discussed in [257, 258]. The dynamics of

the $A + B \to \emptyset$ reaction front was discussed in Zeldovich's early publication [242] and independently rediscovered by Gálfi and Rácz [259]; the steady-state version of the problem is treated in [260].

13.7 Problems

13.1 Fill in the details of solving for the survival probability of a diffusing particle in the absorbing interval $[0, L]$:

(a) First solve the diffusion equation in one dimension, subject to the boundary conditions $c(0, t) = c(L, t) = 0$ and the initial condition $c(x, t = 0) = \delta(x - x_0)$ and thereby obtain the particle concentration given in Eq. (13.9).

(b) Next, integrate over all final points of the particle and average over all starting positions to arrive at the integral expression for the survival probability given in Eq. (13.11).

13.2 Apply the Laplace method to the integral in Eq. (13.11) and show that the long-time asymptotic behavior is indeed given by (13.12).

13.3 Consider a diffusing particle that starts at the origin of an absorbing sphere of radius a. Using spherical symmetry and separation of variables, show that the solution to the radial equation has the form of Eq. (13.17). Compute the amplitudes A_n. Show that, in three dimensions, the Bessel functions can be expressed via trigonometric functions.

13.4 Simulate aggregation in one dimension assuming that the diffusion coefficient varies as the inverse cluster mass, $D_k = 1/k$. Show that the cluster density decays as $t^{-1/3}$ in agreement with heuristic arguments, Eq. (13.46) at $\nu = 1$.

13.5 Compute the integral in Eq. (13.60) and verify the results of Eq. (13.61).

13.6 Use the generating function, $\mathcal{C}(x, z) = \sum_{k \geq 1} c_k(x) z^k$, to solve the steady-state limit of Eq. (13.58).

(a) Show that

$$\mathcal{C}(x, z) = \frac{6D}{Kx_0^2} \left[\frac{1}{(|\xi| + 1)^2} - \frac{1}{(|\xi| + (1 - z)^{-1/3})^2} \right], \qquad \xi = \frac{x}{x_0},$$

with x_0 given by (13.66).

(b) At the position of the source ($x = 0$), the generating function becomes

$$\sum_{k \geq 1} c_k(0) z^k = \frac{6D}{Kx_0^2} \left[1 - (1 - z)^{2/3} \right].$$

Expand the right-hand side and deduce

$$c_k(0) = \frac{4D}{Kx_0^2} \frac{\Gamma(k - 2/3)}{\Gamma(1/3) \, \Gamma(k + 1)} \sim \left(\frac{J^2}{DK} \right)^{1/3} k^{-5/3}.$$

(c) Show that far away from the source, $|x| \gg x_0$,

$$c_k(x) = \frac{12D}{Kx_0^2 |\xi|^3} \frac{\Gamma(k+1/3)}{\Gamma(1/3)\,\Gamma(k+1)} \sim \left(\frac{J^2}{DK}\right)^{1/3} |\xi|^{-3} k^{-2/3}\,.$$

13.7 Consider an aggregation with a spatially localized source. Complete the discussion in the text for a general spatial dimension and show the total number of clusters grows asymptotically according to

$$\mathcal{N} \sim \begin{cases} t, & d > 4, \\ t/\ln t, & d = d^c = 4, \\ t^{(d-2)/2}, & 2 < d < 4, \\ \ln t, & d = d_c = 2, \\ \text{const.}, & d < 2. \end{cases}$$

13.8 Study aggregation in the half-space $x > 0$ with absorption on the plane $x = 0$. (The true spatial dimension is $d = 3$, so that mean-field rate equations hold.)
   (a) Show that the cluster density $N(x, t)$ satisfies the reaction–diffusion equation

$$\frac{\partial N}{\partial t} = D\nabla^2 N - KN^2,$$

   with boundary condition $N(x = 0, t) = 0$ and initial condition $N(x, t = 0) = 1$ for $x > 0$.
   (b) Argue that the solution must have a scaling form and that the proper scaling ansatz is $N(x, t) = (Kt)^{-1} f(\xi)$, with $\xi \equiv x/\sqrt{Dt}$.
   (c) Show that the ansatz from (b) reduces the governing partial differential equation, namely, the reaction–diffusion equation in (a), to an ordinary differential equation

$$f'' + \frac{1}{2}\xi f' + f(1 - f) = 0\,.$$

   (d) Determine the scaling function f numerically.
13.9 Consider two-species annihilation $A + B \to \emptyset$ in the domain $[-L, L]$, with a steady input of A at $x = L$ and input of B at $x = -L$. The input rates of the two species are equal. The appropriate boundary conditions are

$$Dc_A'\big|_{x=L} = -Dc_B'\big|_{x=-L} = -J, \qquad\qquad Dc_A'\big|_{x=-L} = Dc_B'\big|_{x=L} = 0,$$

corresponding to constant fluxes of A at $x = L$ and B and $x = -L$, and reflection at the opposite sides. Starting with the reaction–diffusion equations (13.74) in the steady state, subject to the above boundary conditions, develop an argument analogous to that in the transient case to show that the width of the reaction zone is given by

$$w \sim \left(\frac{D^2}{JK}\right)^{1/3}\,.$$

13.10 Consider again the steady behavior of the reaction–diffusion equations for two-species annihilation $A + B \rightarrow \emptyset$, with a steady input of A at $x = L$ and input of B at $x = -L$.

(a) Consider the combinations $c_{\pm} = c_A \pm c_B$. Show that the solution for c_- that satisfies the boundary conditions is $c_-(x) = jx/D$.

(v) Use the result for c_- to rewrite the equation for c_+ as

$$c_+'' = \frac{k}{2D} \left[c_+^2 - \left(\frac{jx}{D} \right)^2 \right],$$

where the prime denotes differentiation with respect to x and the boundary conditions are $Dc_+' = j$ at $x = L$ and $Dc_+' = -j$ at $x = -L$. Show that for large flux, defined as $j > j_0 = D^2/(KL^3)$, c_+ has the Taylor series expansion

$$\frac{c_+(z)}{c_0} = \left(\frac{4}{5} \right)^{1/3} + \frac{z^2}{(10)^{2/3}} - \frac{z^4}{40} + \cdots,$$

where $z \equiv x/x_0 \ll 1$, $x_0 = (D^2/jK)^{1/3}$, and $c_0 = (j^2/KD)^{1/3}$.

(c) Show that the concentration of the minority species outside the reaction zone is asymptotically given by

$$c_B(x)/c_0 \sim z^{-1/4} \exp\left(-2z^{3/2}/3 \right).$$

To obtain this result, substitute the large-x approximation $c_A \cong c_B + (jx/D)$ in the steady-state equation for c_B to give $Dc_B'' = kc_B(c_B + jx/D)$ and determine the approximate solution of this equation for large x.

Complex networks

The equilibrium and non-equilibrium behaviors of lattice systems constitute a major part of solid state physics, statistical physics, etc. Although the properties of the underlying lattice can affect the behavior of many-particle systems, the lattice primarily provides the framework upon which the physical system is built. If, however, we want to study a dynamical system on complex networks (Fig. 14.1), network structure becomes an essential ingredient of the problem. While one might anticipate that the structure of such networks is a static problem, we will show that a kinetic approach easily gives many statistical characteristics of such complex networks.

14.1 Non-lattice networks

Non-lattice networks were first introduced to provide the scaffolding upon which dynamical systems exhibit simpler equilibrium and non-equilibrium behaviors than lattice systems. Two particularly useful examples of such non-lattice networks are the complete graph and

(a)   (b)
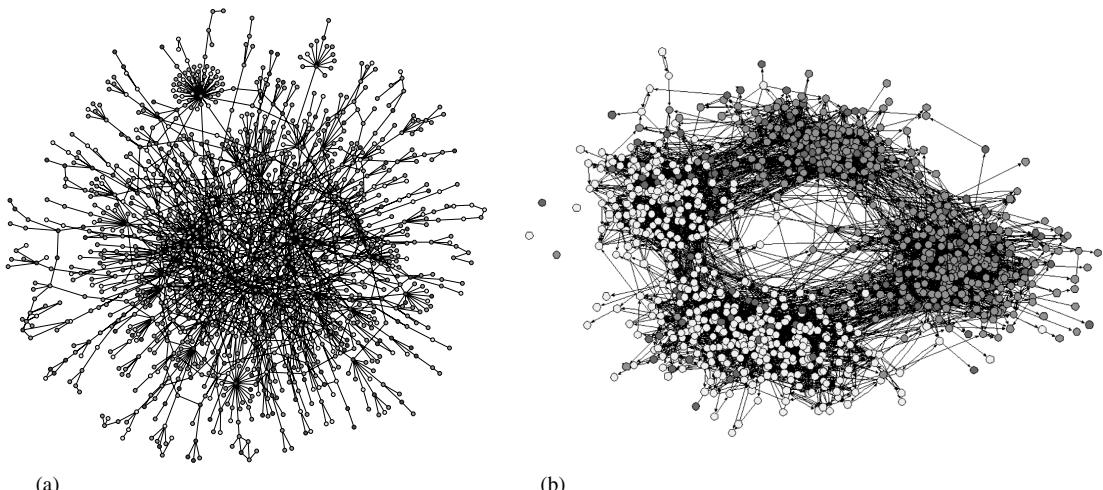
Fig. 14.1. Examples of complex networks: (a) a protein interaction network, reprinted from "Lethality and centrality in protein networks," H. Jeong *et al.*, *Nature* 411, 41 (2001); (b) a representation of high-school social networks, reprinted from "Race, school integration, and friendship segregation in America," J. Moody, *Am. J. Sociol.* 107, 679 (2001).

the Cayley tree. An important feature of these examples is that collective behaviors on these graphs provide the mean-field approximation for cooperativity on lattice systems. However, the complete graph and the Cayley trees are homogeneous, but networks of high complexity are needed to describe, for example, the interaction of neurons in the brain or Internet traffic between interconnected web pages. If we want to study dynamics on such complex networks, we should specify the network completely, that is, list all links between nodes.[1] This task alone is complicated: the structures of various biological networks are at best approximately known, and their large size leads to formidable computational requirements. In addition to these impediments to characterizing structure, we typically want to study a complicated dynamical system on such a network. To help overcome some of these problems, we adopt the statistical physics perspective that precise details are secondary in a large network: if the network is sufficiently large and random, its statistical properties will be close to those properties that characterize a "typical" network.

Example 14.1. *Planar regular random graphs*. Consider a network in which every node is connected with exactly z other nodes and suppose that the network can be embedded in a plane in such a way that its links do not intersect; otherwise the network is arbitrary. Such networks are called planar regular random graphs. There are many such networks, e.g. Cayley trees with coordination number z and two-dimensional lattices (which exist only when $z = 3, 4$, and 6). These examples, however, are special and give a very wrong impression – a *typical* planar regular random graph has *dimension* equal to 4, even though the graph is planar.

We must, of course, explain what we mean by dimension. The simplest definition of a set of dimension D is based on the property that the volume of a ball of radius R grows as R^D. Thus let's count the number of nodes $N(R)$ within a distance $\leq R$ from a given node.[2] The interesting behavior occurs when $R \gg 1$. For lattices, $N \sim R^2$, so planar lattices are indeed two-dimensional. For the Cayley tree, $N \sim (z - 1)^R$; the exponential growth of N with R reflects the (hidden) infinite-dimensional nature of Cayley trees. For the typical planar regular random graph, the asymptotic behavior is strikingly different: $N \sim R^4$, so that their dimension $D = 4$. The vast majority of planar regular random graphs have $D = 4$; only a few exceptional networks, whose probability quickly vanishes with system size, will have a different dimensionality.

The interplay between planarity and the intrinsic $D = 4$ dimension leads to unusual behaviors of various interacting systems defined on planar regular random graphs. For example, the Ising model on such graphs exhibits a very different critical behavior than the Ising model on two-dimensional lattices.

In addition to assuming that a typical network chosen at random from a certain ensemble provides a faithful representation of an actual network, we adopt a dynamical viewpoint and use the tools from non-equilibrium statistical physics to describe various geometrical properties of complex networks. It may seem surprising, at first glance, that this viewpoint

---

[1] In this chapter we use the terminology "nodes" for sites and "links" for bonds.
[2] The distance between the nodes is defined as the number of links along the shortest path that connects these nodes.

is viable and can describe the geometrical properties of a network, particularly when the network is static. However, by viewing a static network as a single snapshot from a suitable evolution process, a dynamical treatment becomes both appropriate and potent. There is a useful parallel with Chapter 7 on adsorption, where the hard-to-calculate final coverage of a surface – an ostensibly static quantity – was derived, in a much simpler way, as a by-product from the solution for the full time-dependent coverage.

Nevertheless, this dynamical perspective is not always feasible or useful. For instance, the global constraint of planarity seems to preclude the possibility of describing regular planar random graphs dynamically.[3] If we dispense with the requirements of planarity and regularity (for example, rather than demanding strict regularity, each node is assumed to be connected to z neighbors on average), the situation drastically simplifies. The resulting random graphs admit a dynamical description in spite of their static definition.

14.2 Evolving random graphs

While we will primarily study the evolution of random graphs, it is helpful to begin with a static definition. In this framework, evolving random graphs are often called Erdős–Rényi (ER) random graphs (Fig. 14.2). There are two generic, and closely related, forms of this model. In the probabilistic version, the graph consists of N nodes, in which the probability that each node pair is connected by a link is p/N, with $0 \leq p \leq N$. The factor $1/N$ is conventionally included because it leads to a percolation transition at $p = 1$ (see below). Since each node has $N - 1$ neighbors, the average number of links attached to a node, also known as the *node degree*, is $p(N - 1)/N$ (and approaches p as $N \to \infty$). In contrast to regular lattices, the degree (or coordination number) of each node is not the same on the ER graph (and indeed for any complex network). Thus the *degree distribution* provides a new characterization of a complex network.
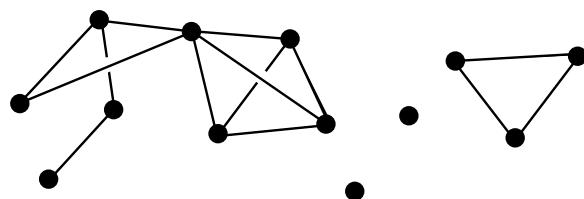
Fig. 14.2.   A realization of the ER graph of 13 nodes and 14 links. The network is partitioned into four clusters: one of size 8, one of size 3, and two of size 1. In this graph there is one node of degree 5, four of degree 3, four of degree 2, one of degree 1, and two of degree 0. In the cluster of size 3 each node has degree 2.

---

[3] Planarity constrains the structure of regular random graphs whenever the coordination number z is at least three. For $z = 1$ and $z = 2$, the planarity is automatically satisfied and a dynamical framework is possible (problem 14.1).

An alternative model is to randomly connect the $N(N-1)/2$ pairs of nodes using a fixed number of links L. Consequently, the fraction of connected links is $2L/[N(N-1)]$. Equating this fraction with p/N, the fixed-link rule is essentially the same, in the limit $N, L \to \infty$, as the probabilistic ER random graph with $p = 2L/N$.

While the ER random graph construction is simple, its geometric properties are amazingly rich. Perhaps the most striking feature is the existence of a percolation transition, as p passes through a critical value $p_c = 1$, in which the nature of the cluster size distribution changes dramatically. A cluster is defined as the maximal set of nodes that are connected by links, so that the graph consists of the union of disjoint clusters (Fig. 14.2). For $p < p_c$, clusters are small and tree-like. For $p > p_c$, there exists a "giant" cluster that consists of a non-zero fraction of all nodes. At $p = p_c$, the cluster size distribution decays algebraically with size. Many of these properties have been elucidated using probabilistic approaches. In this section, we present an alternative, kinetic approach for unraveling the structure of the ER graph.

Kinetic formulation

We recast the ER graph as a kinetic problem by starting with N isolated nodes and introducing links one by one between randomly selected node pairs. These two nodes could be the same and also more than one link may be created between a pair of nodes. However, these two processes occur with a vanishingly small probability, when $N \to \infty$, and may be ignored.[4] For convenience, we set to $N/2$ the rate at which links are introduced. Consequently, the total number of links at time t is $Nt/2$, so that the average degree $2L/N$ equals t. Therefore the average degree grows at rate 1.

Let's first determine the degree distribution of the ER graph. It is convenient to work with the normalized degree distribution n_k, defined as the *fraction* of nodes of degree k. Nodes of degree k are created at rate 1 by introducing a link that attaches to a node of degree $k-1$; similarly nodes of degree k are lost (they become nodes of degree $k+1$) at rate 1 by linking to other nodes. The degree distribution therefore satisfies

$$\frac{dn_k}{dt} = n_{k-1} - n_k .$$
(14.1)

This equation applies for all $k \geq 0$ if we impose the additional condition $n_{-1} \equiv 0$. For the initial condition $n_k(0) = \delta_{k,0}$, corresponding to N initially disconnected nodes, the solution to Eq. (14.1) is the Poisson distribution (see problem 14.2)

$$n_k = \frac{t^k}{k!} e^{-t} .$$
(14.2)

From this solution we find that the mean degree is indeed equal to the time, $\langle k \rangle = \sum k n_k = t$, while the standard deviation is $\sqrt{\langle k^2 \rangle - \langle k \rangle^2} = \sqrt{t}$.

---

[4] Below, we will describe the creation of cycles where the possibility of creating multiple links and loops will be taken into account.

We now investigate the time evolution of the cluster size distribution, from which we can probe the percolation transition of the ER graph. Initially the network consists of N isolated single-site clusters. As links are introduced into the network, clusters can only merge, so that the number of clusters systematically decreases and their mean size grows. The addition of a new link will lead to merging of two clusters. The probability that two disconnected clusters of sizes i and j join to create a cluster of size $k = i + j$ equals $(i\, C_i/N) \times (j\, C_j/N)$; here C_j is the total number of clusters of size j. Following the same reasoning that led to Eq. (5.34) in Section 5.2, the master equation for $c_k(t) = C_k(t)/N$, the density of clusters with k nodes at time t, is

$$\frac{dc_k}{dt} = \frac{1}{2} \sum_{i+j=k} i c_i\, j c_j - k\, c_k. \tag{14.3}$$

This equation coincides with that for the cluster mass distribution in product-kernel aggregation. Since the network starts with all nodes isolated, the initial condition is $c_k(0) = \delta_{k,1}$. From Eq. (5.44), the cluster size distribution is

$$c_k(t) = \frac{k^{k-2}}{k!}\, t^{k-1}\, e^{-kt}\,. \tag{14.4}$$

We now exploit this basic result, as well as several other relevant results from Section 5.2, to determine a variety of geometrical features of the ER graph.

The percolation transition

The first basic consequence of Eq. (14.4) is that there is a transition as t passes through 1. Applying Stirling's approximation to this equation, the asymptotic behavior of c_k is given by (see also Eq. (5.45))

$$c_k(t) \simeq \frac{1}{\sqrt{2\pi}\, k^{5/2}}\, e^{-k(t - \ln t - 1)} \simeq \frac{1}{\sqrt{2\pi}\, k^{5/2}}\, e^{-k(1-t)^2/2}\,, \tag{14.5}$$

as $t \uparrow 1$. For $t < 1$, the cluster size distribution c_k decays exponentially with k, so that the mean cluster size is finite. Naively, the mean size would seem to be given by $\sum k\, c_k$. By construction, however, this sum equals the total mass density, which equals one. A more appropriate measure of the mean cluster size is the second moment, $M_2 \equiv \sum_k k^2 c_k$. At $t = 1$, the cluster size distribution has a power-law tail with exponent $-5/2$, and this implies that the mean cluster size diverges in an infinite network. Therefore a percolation transition occurs at $t = 1$.

For an infinite ER graph, the singular behavior near the percolation transition is best appreciated by studying the first few moments of the cluster size distribution, $M_n \equiv \sum_k k^n c_k$. It is important to appreciate (see Section 5.3 and specifically the discussion on page 145) that the sum does not include the contribution of the giant component, or more informally,

the infinite cluster, whenever it happens to exist. For the zeroth moment, we have

$$
M_0 = \begin{cases}
1 - t/2, & t \le 1, \\
1 - t/2 + 2(t-1)^3/3 + \cdots, & t \downarrow 1, \\
e^{-t} + (t/2)\, e^{-2t} + \cdots, & t \to \infty.
\end{cases}
\tag{14.6}
$$

A decisive signal of the percolation transition is provided by the *gel fraction* $g \equiv 1 - M_1$, namely, the fraction of nodes that are part of the giant cluster. The evolution for the gel fraction is implicitly determined by $g = 1 - e^{-gt}$ (Eq. (5.47)), and the limiting behaviors of this gel fraction are given by

$$
g = \begin{cases}
0, & t < 1, \\
2(t-1) - 8(t-1)^2/3 + \cdots, & t \downarrow 1, \\
1 - e^{-t} - t e^{-2t} + \cdots, & t \to \infty.
\end{cases}
\tag{14.7}
$$

Finally, for the initial condition $M_2(0) = 1$, the second moment is given by

$$
M_2(t) = \begin{cases}
(1-t)^{-1}, & t < 1, \\
(e^{gt} - t)^{-1}, & t > 1.
\end{cases}
\tag{14.8}
$$

Thus an infinite cluster forms at $t = 1$ on an infinite ER graph. As t increases far beyond the percolation point, the mean size of the remaining finite clusters goes to zero.

Since one normally studies a large but finite ER graph, finite-size scaling is an integral part of the description of the percolation transition. To appreciate the role of finite-size effects, let's first determine the size M^* of the largest cluster in a finite evolving random graph. This size may be estimated by the extremal criterion (see the highlight on page 17)

$$
N \sum_{k \ge M^*} c_k = 1,
\tag{14.9}
$$

which states that there will be a single cluster whose size is $\ge M^*$. Using the asymptotic forms in Eqs (14.5) for $c_k(t)$ and approximating the above sum by an integral, the extremal criterion gives

$$
M^* \sim \begin{cases}
\ln N / (1-t)^2, & t < 1, \\
N^{2/3}, & t = 1, \\
gN, & t > 1.
\end{cases}
\tag{14.10}
$$

For a finite network, the sum for the mean cluster size should be cut off at the largest cluster size of M^*, so that

$$
M_2 \sim \sum_{k < M^*} k^2 c_k \sim \int^{M^*} k^2\, k^{-5/2}\, dk \sim \sqrt{M^*} \sim N^{1/3}.
$$

For a finite ER graph, the mean cluster size $M_2 = (1 - t)^{-1}$ will be truncated when it becomes of the order of $N^{1/3}$. Consequently, the sharp percolation point becomes smeared into a critical regime, also known as the *scaling window*, that is defined by

$$|1 - t| \sim N^{-1/3}. \tag{14.11}$$

In addition to the percolation transition where the giant cluster is born, there is a second transition at a later time t_1 that demarcates the regime where the graph consists of many clusters and a regime when the entire network becomes connected into a single cluster. That is, at t_1 the giant cluster engulfs the entire network and no finite clusters remain. Using the criterion $NM_0(t_1) \sim 1$ and Eq. (14.6), we find $t_1 \simeq \ln N$.

Geometrical properties

We can understand the geometrical properties of a large ER graph by exploiting the logical consequences of the kinetic formulation and by using probabilistic arguments. First consider the regime below the percolation threshold. Since there are of the order of N small clusters for $t < 1$ (Eq. (14.6)), a newly introduced link will typically appear between two disjoint clusters rather than between two nodes in the same cluster. Consequently, clusters are predominantly trees.

Above the percolation threshold, the size of the largest cluster equals gN, but the size of the second-largest cluster is governed by the same exponential decay in c_k as in the largest cluster below percolation. Thus the size of this second-largest cluster scales as $(t-1)^{-2} \ln N$. When the average node degree p is larger than, but of the order of, one, the structure of the ER graph just above percolation is roughly that of a branching tree. We can exploit this picture to give a simple estimate of the diameter of the network. For $p \gtrsim 1$, an arbitrary node is connected, on average, to p first-generation nodes; each first-generation node is connected to $p - 1$ second-generation nodes, etc.; and each $(k - 1)$st-generation node is connected to $p - 1$ nodes in generation k. Because $p/N \ll 1$, the probability of forming closed loops is of the order of $1/N$ in early generations of this branching tree. Thus the network diameter increases by one each time a new generation is added. However, when the total number of nodes $p(p - 1)^k$ in a tree of k generations becomes of the order of N, then additional generations must lead to closed loops and the diameter saturates at this point. Thus the diameter of the ER graph is of the order of $\ln N / \ln p$. This logarithmic dependence of diameter on N is a generic feature of many complex networks.

The ER graph exhibits some unusual features due to the influence of long-range connections that cannot occur for percolation on a regular lattice. Perhaps the most striking is the existence of a sequence of thresholds for the disappearance of all clusters of a given finite size. Let's start by determining the point where monomers can no longer exist. From the expression for c_k in Eq. (14.4), the total number of monomers C_1 is given by $C_1 = Ne^{-t}$. Thus a single monomer typically remains up to time $t_1 = \ln N$, as found previously by the criterion that the network becomes completely connected. In the same spirit, the number of k-mers $C_k = Nc_k$ first increases, but later decreases back to 1 when $t_k \simeq (1/k) \ln N$. Thus there exists a series of threshold times, given by $t_k \simeq (1/k) \ln N$, where the last k-mer

disappears. Within the time interval (t_2, t_1), the network contains a giant component and isolated nodes (monomers); therefore monomers are the last type of small components to disappear from the network.

Cycles

We can generally distinguish three types of clusters: *tree*, *unicyclic*, or *complex*. By definition, a tree contains no closed paths (cycles), a unicyclic cluster contains a single cycle, while a complex cluster contains at least two cycles. Alternatively, we can describe these types of clusters using the notion of Euler characteristic. For a graph, the Euler characteristics is defined as the total number of links minus the total number of nodes. The Euler characteristic of any tree is -1; for a unicyclic cluster, the Euler characteristic equals zero; a complex cluster has a positive Euler characteristic. For example, the graph in Fig. 14.2 contains two trees (isolated nodes), a unicyclic cluster of size 3, and a complex cluster of size 8 (whose Euler characteristic is $11 - 8 = 3$).

An ER graph starts with N disconnected nodes, i.e. N trees of size 1. Adding a link between two trees gives rise to a larger tree. During the initial stage of evolution, the random graph is a *forest*; that is, all its clusters are trees. Eventually a link inside a tree will be added, and this tree will become a unicyclic component. Adding a link between a tree and a unicyclic cluster results in a unicyclic cluster, while a link between two unicyclic clusters leads to a complex cluster. Because there will be a spectrum of clusters of various complexities, it would seem that the classification into trees, unicyclic clusters, and complex clusters is insufficient. Instead we should distinguish complex clusters more precisely and group them according to their Euler characteristic. Remarkably, this precision is unnecessary for large networks ($N \to \infty$), since there is at most one complex cluster. More precisely,

1. below the percolation threshold, there is a macroscopic ($\sim N$) number of trees and a finite number (fluctuating from realization to realization) of unicyclic clusters, and
2. above the percolation threshold, the situation is the same as below the threshold, except there is a single complex cluster, the giant component.

Let's therefore determine the rate at which clusters of a given type are created. For trees, this rate is proportional to $N^{-1} \times N^2$, where N^{-1} is the rate to add a link to a node and N^2 counts the number of pairs of trees. Unicyclic clusters are created at a rate proportional to $N^{-1} \times N$, since they arise when the link inside a tree is added. Clusters of complexity 1 are generated at a rate proportional to N^{-1}, and therefore the number of complex clusters vanishes in the $N \to \infty$ limit. The only exception to this rule is the giant component, which is complex due to the continuous addition of links inside it. Thus we conclude that apart from the giant component there are no complex clusters. We also see that, among the C_k clusters of size k, all are trees except for a finite number U_k of unicyclic clusters.

In principle, both C_k and U_k are random quantities that fluctuate from realization to realization. We know, however, that $C_k \sim N$, while $U_k \sim 1$. This scaling suggests that the former quantity has relatively small fluctuations,

$$C_k = N c_k + \sqrt{N}\, \xi_k \,, \tag{14.12}$$

where c_k is a deterministic density that satisfies the master equations (14.3). The evolution of the random quantities ξ_k can, in principle, be analyzed using the techniques described in Chapter 12; technically this is cumbersome, and given that fluctuations are small compared to the average, such a calculation provides little additional insights.

Since the number of unicyclic clusters is finite, this quantity strongly varies between different realizations of the network, so that its average $u_k = \langle U_k \rangle$ provides limited information. Unfortunately, it is difficult to compute the distribution $P_k(U_k)$. Furthermore, the single-variable distributions $P_k(U_k)$ are not very useful, as the quantities U_k are highly correlated ($\langle U_i U_j \rangle \neq \langle U_i \rangle \langle U_j \rangle$). Hence the proper characterization is provided by the probability for the number of unicyclic clusters of all sizes, i.e. $P(\mathbf{U}, t)$, where $\mathbf{U} = \{U_k\}$.

What we can do is to determine the average number u_k of unicyclic clusters of different sizes. The governing equations for the u_k are (compare with Eq. (14.3))

$$\frac{du_k}{dt} = \frac{1}{2} k^2 c_k + \sum_{i+j=k} i u_i \, j c_j - k \, u_k. \tag{14.13}$$

The first term on the right describes the creation of a unicyclic cluster by adding a link inside a tree;[5] the next two terms account for merging of two different clusters, one of which is unicyclic. Before solving (14.13), we first determine the average number of unicyclic components $U = \sum u_k$ in the sol phase ($t < t_g = 1$). Summing Eqs (14.13) and using (14.8) we obtain

$$\frac{dU}{dt} = \frac{1}{2} M_2 = \frac{1}{2(1-t)}, \tag{14.14}$$

from which

$$U(t) = \frac{1}{2} \ln \frac{1}{1-t}. \tag{14.15}$$

The total number of unicyclic components remains finite in the sol phase and diverges at the percolation point. For a finite system, the percolation point broadens into the scaling window (14.11), where $1 - t \sim N^{-1/3}$. Therefore the total number of unicyclic clusters diverges logarithmically in the critical regime

$$U_g = \frac{1}{6} \ln N. \tag{14.16}$$

The natural way to solve for the u_k is to recast the infinite set of equations (14.13) into a partial differential equation for the generating function, a strategy that proved extremely effective in dealing with the master equations for aggregation in Chapter 5. Here we use a slight variation of the method that gives the solution with less work. The idea is to infer the general time dependence by solving Eqs (14.13) for the first few k, subject to $u_k(0) = 0$.

---

[5] Recall that according to our definition of the network evolution, a node can link to itself and two nodes can be connected by more than one link (thereby creating a multigraph).

We thereby find

$$u_1 = \frac{1}{2} t e^{-t}, \quad u_2 = \frac{3}{4} t^2 e^{-2t}, \quad u_3 = \frac{17}{12} t^3 e^{-3t}, \quad \text{etc.,}$$

suggesting that $u_k \sim t^k e^{-kt}$. Hence we seek a solution in the form[6]

$$k u_k(t) = \frac{1}{2} A_k t^k e^{-kt} . \tag{14.17}$$

Substituting (14.4) and (14.17) into (14.13) we obtain

$$A_k = \frac{k^k}{k!} + \sum_{i+j=k} A_i \frac{j^{j-1}}{j!} . \tag{14.18}$$

This recurrence can be solved by recasting it into an ordinary differential equation[7] for the generating function $\mathcal{A}(z) = \sum A_k e^{kz}$. From this approach, we obtain the prefactors $A_k = \sum_{0 \le n \le k-1} k^n/n!$ (problem 14.4), so that

$$u_k(t) = \frac{1}{2} t^k e^{-kt} \sum_{0 \le n \le k-1} \frac{k^{n-1}}{n!} . \tag{14.19}$$

The sum approaches $(k^k/k!)\sqrt{\pi/(2k)}$ in the large-k limit (problem 14.5). Therefore

$$u_k(t) \simeq \sqrt{\frac{\pi}{8k}} \frac{(kt)^k}{k!} e^{-kt} \quad \text{when} \quad k \gg 1. \tag{14.20}$$

Exactly at the percolation point the average size distribution of the unicyclic clusters varies inversely proportional to the size,

$$U_k(t_g) = \frac{1}{2} e^{-k} \sum_{0 \le n \le k-1} \frac{k^{n-1}}{n!} \simeq \frac{1}{4k} . \tag{14.21}$$

Owing to the large fluctuations in u_k, its average gives limited information about cycles. However, the methods we have used provide useful probabilistic information. Let us look at one such property, the probability distribution that a certain number of cycles exists.

Example 14.2. *No cycles.* What is the probability $s_0(t)$ that the system does not contain a cycle up to time t? It suffices to answer this question in the pre-gel regime, as the giant

---

[6] Since the quantity $k u_k$ appears on the right-hand side of (14.13), it is natural to write the ansatz (14.17) in terms of $k u_k$.

[7] This ordinary differential equation is easier to solve than the partial differential equation for the generating function $\sum u_k(t) e^{kz}$. The trick of extracting the time dependence applies to many other examples treated in this book. For example, for the cluster size distribution $c_k(t)$ for the evolving random graph, it is easy to guess that $c_k \sim t^{k-1} e^{-kt}$. Then the k-dependent prefactor and thereby the full solution (14.4) can be found with somewhat less effort than the original generating function solution. The disadvantage of this method is that it applies only to special initial conditions.

component certainly contains cycles. From (14.14), the production rate of cycles is $J = 1/[2(1-t)]$. Because the production of cycles is a random process, the probability of having no cycles obeys

$$\frac{ds_0}{dt} = -Js_0 \,, \tag{14.22}$$

with the initial condition $s_0(0) = 1$. The solution is $s_0(t) = \sqrt{1-t}$ for $t \le 1$. At the percolation point, we estimate the probability for no cycles to exist by using the width of the critical region $N^{-1/3}$ for $|1-t|$ (Eq. (14.11)). Therefore the probability of having no cycles is $s_0 \sim \sqrt{N^{-1/3}} = N^{-1/6}$.

Similarly we can compute the probability $s_n(t)$ of having exactly n unicyclic clusters.[8] Given that the production of cycles is a random process, the probabilities $s_n(t)$ obey

$$\frac{ds_n}{dt} = J(s_{n-1} - s_n).$$

Solving these equations subject to the initial condition $s_n(0) = \delta_{n,0}$ we find that the solution is the Poisson distribution

$$s_n = \frac{(1-t)^{1/2}}{n!} \left[\frac{1}{2} \ln \frac{1}{1-t} \right]^n. \tag{14.23}$$

14.3 Random recursive trees

A prototypical and illustrative example of a growing network is the *random recursive tree* (RRT). In this network, nodes are added one by one, and each new node attaches to a single "target" node that is chosen randomly among the existing nodes. By the restriction that each new node has a single point of attachment, the resulting network has a tree structure,[9] a feature that simplifies the analysis of the system. The growth rules of the RRT thus are as follows (and see Fig. 14.3):

1. Pick one of the nodes of the RRT – the target – with uniform probability.
2. Introduce a new node that links to the target node.

Starting with an initial state that consists of a single node, these two steps are repeated until the tree reaches a desired number of nodes N. The RRT is a particularly useful model because many of its geometrical properties are readily soluble. Moreover, there are a variety of physically motivated and simple ways to generalize the RRT that lead to rich network structures.

---

[8] Here we disregard more subtle characteristics like the size of these clusters, the length of cycles, etc.
[9] If a new node attaches to more than one pre-existing node, closed loops can form. The degree distribution of such a network is modified only slightly compared to growing trees, but other features, such as the network diameter, are strongly influenced by the existence of loops.
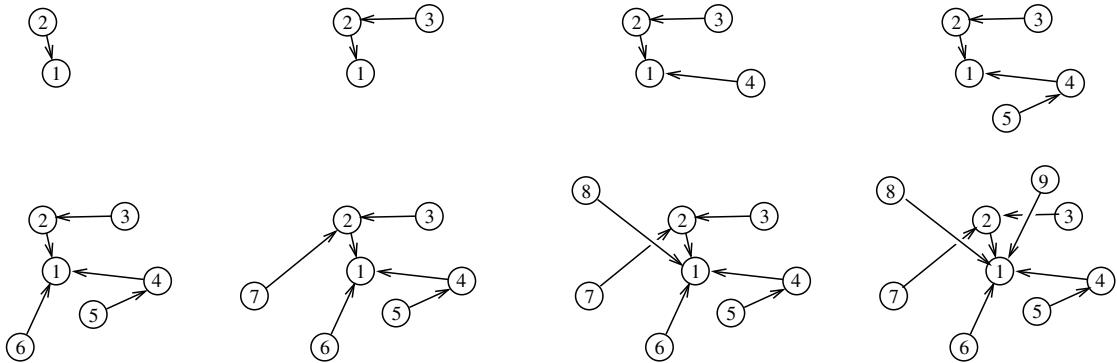
Fig. 14.3. The first few evolution steps of a random recursive tree.

The degree distribution

A basic local characterization of the RRT – and indeed any complex network – is its degree distribution. We first outline the steps needed to derive the exact degree distribution (see also problem 14.7) and then turn to an asymptotic analysis of the degree distribution in the limit of large N. The degree state of any network is characterized by the vector $\mathbf{N} \equiv \{N_1, N_2, \ldots\}$, where N_k denotes the number of nodes of degree k. For the case of the RRT, when a new node is introduced, the network state vector $\mathbf{N}$ evolves as follows:

$$\begin{aligned} \text{attach to node of degree 1} \quad & (N_1, N_2) \rightarrow (N_1, N_2 + 1), \\ \text{attach to node of degree } k > 1 \ & (N_1, N_k, N_{k+1}) \rightarrow (N_1 + 1, N_k - 1, N_{k+1} + 1). \end{aligned} \tag{14.24}$$

While one can in principle write the evolution equation for the full probability distribution $P(\mathbf{N})$, this equation provides "too much" information. Typically we are interested in the average number of nodes of a given degree, $\langle N_k \rangle$ (the degree distribution), or perhaps the two-body correlations functions $\langle N_i N_j \rangle$. Here the angle brackets denote an average over all possible growth histories of the network. As we now discuss, the master equation approach is ideally suited to determine such quantities.

As indicated by Eq. (14.24), we need to separately consider nodes of degree 1 and nodes of degree greater than 1. The number of nodes of degree 1, $N_1(N)$, is a random variable that changes according to

$$N_1(N+1) = \begin{cases} N_1(N), & \text{probability} \quad N_1/N, \\ N_1(N) + 1, & \text{probability} \quad 1 - N_1/N, \end{cases} \tag{14.25}$$

after the addition of each new node. Namely, with probability N_1/N, the new node attaches to a node of degree 1 and the number of such nodes does not change. Conversely, with probability $1 - N_1/N$, the new node attaches to a node of degree $k > 1$ and N_1 thus increases by 1. The resulting evolution equation for the average number of nodes of degree 1

is therefore given by

$$\langle N_1(N+1)\rangle = \left\langle N_1(N) \times \frac{N_1(N)}{N}\right\rangle + \left\langle (N_1(N)+1) \times \left(1 - \frac{N_1}{N}\right)\right\rangle$$

$$= 1 + \left(1 - \frac{1}{N}\right)\langle N_1(N)\rangle. \tag{14.26}$$

It is possible to solve this equation and find the average number of nodes of degree 1. The answer is

$$\langle N_1(N)\rangle = \frac{N}{2} + \frac{1}{2(N-1)} \tag{14.27}$$

for $N \geq 2$. This result can be obtained by the generating function approach (problem 14.7), and can be readily verified by direct substitution into (14.26).

By similar reasoning, the number of nodes of degree $k \geq 2$, $N_k \equiv N_k(N)$, evolves according to

$$N_k(N+1) = \begin{cases} N_k - 1, & \text{probability} \quad N_k/N, \\ N_k + 1, & \text{probability} \quad N_{k-1}/N, \\ N_k, & \text{probability} \quad 1 - (N_{k-1} + N_k)/N, \end{cases} \tag{14.28}$$

after each node addition. Following the same steps that led to Eq. (14.26), $\langle N_k\rangle$ evolves by

$$\langle N_k(N+1)\rangle = \langle N_k(N)\rangle + \left\langle \frac{N_{k-1}(N) - N_k(N)}{N}\right\rangle. \tag{14.29}$$

Equation (14.29) can again be solved to give the exact degree distribution for finite networks (problem 14.7).

In what follows, we restrict ourselves to the much simpler leading behavior of the degree distribution in the limit $N \to \infty$. To minimize notational clutter, we drop the angle brackets and write N_k for the average number of nodes of degree k in a network that consists of N total nodes. Next we replace the discrete differences with derivatives in Eqs (14.26) and (14.29), so that the asymptotic degree distribution evolves according to

$$\frac{dN_k}{dN} = \frac{N_{k-1} - N_k}{N} + \delta_{k1} . \tag{14.30}$$

This equation is similar to that for the ER graph, Eq. (14.1), except for the additional delta-function term that accounts for the introduction of the new node of degree 1. To get a feeling for the nature of the solution, let's solve Eqs (14.30) one by one. The equation for N_1 is $dN_1/dN = -N_1/N + 1$, with solution $N_1 = N/2$. Then the equation for N_2 is $dN_2/dN = (N_1 - N_2)/N$, with solution $N_2 = N/4$. Continuing this approach shows that

all the N_k are proportional to N. It therefore is convenient to work with the density of nodes of degree k, $n_k \equiv N_k/N$, in terms of which Eq. (14.30) reduces to

$$n_k = n_{k-1} - n_k + \delta_{k1}. \tag{14.31}$$

Starting with $n_0 = 0$, we obtain $n_1 = \frac{1}{2}$, $n_2 = \frac{1}{4}$, etc., and the general solution is $n_k = 2^{-k}$. From this distribution, the average degree $\langle k \rangle = 2$, while the largest degree k_{max}, which follows from the extremal criterion $\sum_{k>k_{max}} N\, 2^{-k} = 1$ (see the highlight on page 17), gives $k_{max} \simeq \ln N / \ln 2$. Because the degree distribution of the RRT is rapidly decaying with k, even the largest degree is, in some sense, small.

Redirection

We now generalize the RRT to incorporate the mechanism of *redirection*. In redirection, the network is constructed by repeatedly applying the following steps (Fig. 14.4):

1. Pick a pre-existing node x from the network uniformly at random.
2. With probability $1 - r$, the new node n attaches to x.
3. With probability r, node n attaches to an *ancestor* node y of node x.[10]

An important motivation for this construction is that redirection unexpectedly leads to *linear preferential attachment,* a fundamental growth mechanism for complex networks that will be discussed in detail in the next section.[11]
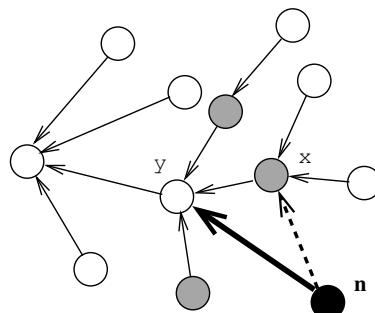
Fig. 14.4. Illustration of redirection. The rate at which attachment occurs to the ancestor **y** of node **x** by redirection is proportional to the number of upstream neighbors to **y** (shaded).

[10] There is a technical subtlety because redirection requires that every node has an ancestor. To ensure this condition always holds, the initial state should consist of at least two nodes and one link, with each node defined as the ancestor of the other.

[11] Since every node addition requires just two elemental instructions, redirection provides an extremely efficient way to simulate very large networks that grow by preferential attachment.

According to redirection, the degree distribution now evolves according to

$$\frac{dN_k}{dN} = \frac{1-r}{N}\left[N_{k-1} - N_k\right] + \delta_{k1} + \frac{r}{N}\left[(k-2)N_{k-1} - (k-1)N_k\right]$$

$$= \frac{r}{N}\left\{\left[k-1+\left(\frac{1}{r}-2\right)\right]N_{k-1} - \left[k+\left(\frac{1}{r}-2\right)\right]N_k\right\} + \delta_{k1}, \qquad (14.32)$$

where the second line is obtained by a straightforward rearrangement of terms. In the first line, the first three terms correspond to the RRT, whose evolution equation (14.30) is recovered for redirection probability $r = 0$. The last two terms account for the change in N_k due to redirection. To understand their origin, consider the gain term. Since the initial node is chosen uniformly, if redirection does occur, then the probability that a node of degree $k - 1$ receives the newly redirected link is proportional to the number of its upstream neighbors (shaded nodes in Fig. 14.4), which equals $k - 2$. A similar argument applies for the redirection-driven loss term. Thus uniform attachment, in conjunction with redirection, generates linear preferential attachment because the rate at which the new node attaches to an existing node of degree k is a linear function of k, $A_k = k + (1/r - 2)$. In the next section, we will study preferential attachment networks in a systematic way, from which the degree distribution in redirection follows easily.

Genealogical tree and the diameter

A revealing feature of the structure of the RRT is its underlying genealogy. To define this genealogy, we start by defining the initial node to be in generation $g = 0$. Nodes that attach to those in generation g form generation $g+1$, irrespective of when attachment occurs during the growth of the network. For example, in the final network of Fig. 14.3 (reproduced in Fig. 14.5), node 1 is the ancestor of 2, while nodes 3 and 7 are the descendants of 2. There are five nodes in generation $g = 1$ and three in $g = 2$, leading to the genealogy on the right of Fig. 14.5.

How many generations are contained in a tree of N nodes? What is the size $L_g(N)$ of the gth generation? To answer these questions, note that $L_g(N)$ increases by 1 when a new node attaches to a node in generation $g - 1$. When a new node attaches to any pre-existing node
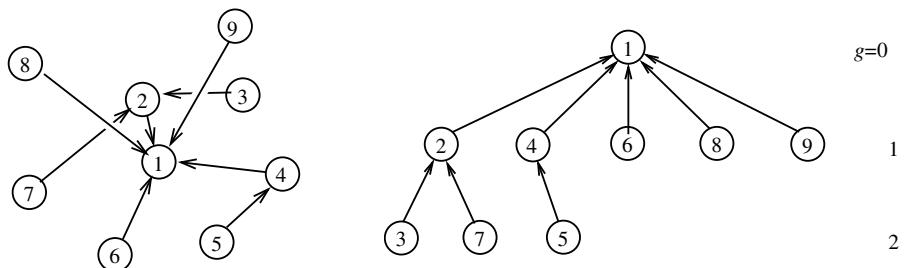
Fig. 14.5. A random recursive tree of nine nodes (from Fig. 14.3) and its corresponding genealogical tree.

equiprobably, attachment to a node in generation $g - 1$ occurs with probability L_{g-1}/N. Consequently L_g evolves as

$$\frac{dL_g}{dN} = \frac{L_{g-1}}{N}, \tag{14.33}$$

with solution $L_g(\tau) = \tau^g/g!$, where $\tau = \ln N$. Using Stirling's approximation, $L_g(N)$ initially grows with g for $g < \tau$, and then decreases and becomes of order 1 when $g = e\,\tau$. The genealogical tree therefore contains $e\,\tau$ generations for a tree of N nodes. Since the diameter is twice the distance from the root of the genealogical tree to the last generation (also the maximum distance between any pair of nodes), the diameter scales as $2e\,\tau \simeq 2e \ln N$.

14.4 Preferential attachment

In *preferential attachment*, the rate A_k at which a node attaches to a pre-existing node of degree k is an increasing function of k. Such a rule encapsulates the notion of the "rich get richer," in which being advantaged now confers the benefit of being able to gain advantage at a higher rate. As an example in the context of scientific citations, preferential attachment corresponds to the feature that a publication with a large number of citations continues to be well cited in the future merely by virtue of being well-cited now. A ubiquitous feature of preferential attachment is that the degree distribution of networks that are built by this mechanism have a broad tail. This feature runs counter to traditional network models, such as the ER random graph discussed in Section 14.2, which have a much steeper Poisson degree distribution.

Master equation

The evolution of the degree distribution for a network whose growth is governed by an attachment rate A_k is (compare with Eq. (14.30) for the RRT)

$$\frac{dN_k}{dN} = \frac{A_{k-1}N_{k-1} - A_k N_k}{A} + \delta_{k1}. \tag{14.34}$$

The first term on the right accounts for the new node connecting to a node that already has $k - 1$ links, thereby increasing N_k by one. Since there are N_{k-1} nodes of degree $k - 1$, the total rate at which such processes occur is proportional to $A_{k-1}N_{k-1}$. The factor $A(N) \equiv \sum_{j \geq 1} A_j N_j$ in the denominator ensures the correct normalization. A corresponding role is played by the second term on the right-hand side. The overall amplitude of A_k is immaterial, since only the ratio A_k/A appears in the master equation. The last term accounts for the introduction of the new node that has one outgoing link and no incoming links. To solve for the degree distribution of such networks, we need to specify the attachment rate A_k. Here we focus on the general case of $A_k = k^\gamma$, with $\gamma \geq 0$. As we shall see, fundamentally

different behaviors arise for sublinear ($\gamma < 1$), superlinear ($\gamma > 1$), and linear ($\gamma = 1$) attachment rates.

Linear, logarithmic and optimal algorithms

How does one actually construct a preferential attachment network? For linear preferential attachment, the probability that a new node attaches to an existing node α with degree k_α is given by $k_\alpha / \sum_\beta k_\beta$. Thus we need to devise a method to select a node α with this probability. Three natural possibilities are as follows:

Linear algorithm. First, form the sequence of cumulative probabilities $\{P_m\}$, with $P_m = \sum_{1 \leq \alpha \leq m} p_\alpha$ for $m = 1, 2, \ldots, N$ (by definition $P_N = 1$). Next, pick a random number r that is uniformly distributed in $[0, 1]$. Finally, attach to the node that satisfies the condition $r > P_m$ but $r < P_{m+1}$ by linearly scanning through the P_m. This linear (and inefficient) algorithm is equivalent to searching for the telephone number of United Airlines in a telephone book (remember those?) by starting at the beginning and scrolling page by page until United Airlines is reached.

Logarithmic algorithm. In computer science courses on algorithms, one learns that linearly searching through a list is unnecessarily inefficient. A variety of methods have been developed to search a list in a time that scales as the logarithm of its length. For example, when we search in a telephone book, we are implicitly performing logarithmic search: thus in looking for United Airlines, we typically first go to the page for the letter "U," then look for the page that contains "Un," then "Uni," etc., until we're on the correct page. This method allows one to find the target node in a time that scales linearly with the network size.

Optimal algorithm. Using the redirection mechanism of the previous section, the task of finding the target node requires of the order of a single operation, *independent* of the size of the network. This $\mathcal{O}(1)$ search is analogous to immediately going to the entry for United Airlines in any telephone book, whether for a small village or for New York City.

Moments and the degree distribution

As a first step to solve for the degree distribution, it is instructive to study the moments $M_\alpha(N) \equiv \sum_j j^\alpha N_j$. This same preliminary step proved useful in understanding the kinetics of aggregation, as discussed in Section 5.2 (see page 137). The zeroth and first moments of this distribution have a simple evolution with N:

$$\frac{dM_0}{dN} = \sum_j \frac{dN_j}{dN} = 1, \qquad \frac{dM_1}{dN} = \sum_j j \frac{dN_j}{dN} = 2. \qquad (14.35)$$

The equation for M_0 states that the total number of nodes of any degree increases by 1 each time a new node is introduced. Similarly, the equation for M_1 states that the total degree of the network, $\sum j N_j$, increases by two upon the addition of a single link. This latter equation arises because the degree of each node at the end of the newly added link increases by one. Thus both the zeroth and first moments of the degree distribution increase linearly with N. For the attachment rate $A_k = k^\gamma$ with $0 \leq \gamma \leq 1$, the total rate $A = \sum_j j^\gamma N_j$ also grows linearly with N, because A is intermediate between the zeroth and first moments.

Thus asymptotically $A \simeq \mu N$, with the as yet undetermined amplitude μ varying smoothly between 1 and 2 as γ increases from 0 to 1 (problem 14.9).

Solving for the first few N_k from Eq. (14.34), it becomes clear that each N_k is also proportional to N (problem 14.10). Thus substituting $N_k(N) = n_k N$ and $A \simeq \mu N$ into the master equations, the overall N dependence cancels, leaving behind the recursion relations

$$n_k = \frac{A_{k-1} n_{k-1} - A_k n_k}{\mu}, \quad k > 1, \quad \text{and} \quad n_1 = -\frac{A_1 n_1}{\mu} + 1,$$

whose formal solution is

$$n_k = \frac{\mu}{A_k} \prod_{1 \leq j \leq k} \left(1 + \frac{\mu}{A_j} \right)^{-1}. \tag{14.36}$$

To make this solution explicit, we need the amplitude μ in $A(N) = \mu N$. Using the definition $\mu = \sum_{j \geq 1} A_j n_j$ and (14.36) we get

$$\sum_{k \geq 1} \prod_{1 \leq j \leq k} \left(1 + \frac{\mu}{A_j} \right)^{-1} = 1. \tag{14.37}$$

Thus the amplitude μ depends, in general, on the functional form of the attachment rate. Equations (14.36) and (14.37) represent the formal solution for the degree distribution of preferential attachment networks. To extract the physical meaning of this solution, we examine its asymptotic behavior for different values of the exponent γ in the attachment rate.

Example 14.3. *Sublinear attachment rate.* For $A_k = k^\gamma$ with $\gamma < 1$, we rewrite the product in Eq. (14.36) as the exponential of a sum, convert the sum to an integral, and then expand the logarithm inside the integral in a Taylor series. These straightforward steps lead to

$$n_k \sim \begin{cases} k^{-\gamma} \exp\left[-\mu \left(\dfrac{k^{1-\gamma} - 2^{1-\gamma}}{1 - \gamma} \right)\right], & \frac{1}{2} < \gamma < 1, \\[2ex] k^{(\mu^2-1)/2} \exp\left[-2\mu \sqrt{k}\right], & \gamma = \frac{1}{2}, \\[2ex] k^{-\gamma} \exp\left[-\mu \dfrac{k^{1-\gamma}}{1 - \gamma} + \dfrac{\mu^2}{2} \dfrac{k^{1-2\gamma}}{1 - 2\gamma}\right], & \frac{1}{3} < \gamma < \frac{1}{2}, \end{cases} \tag{14.38}$$

etc. Each time γ decreases through $1/m$, where m is an integer, an additional term that is an increasing function of k is generated in the exponential. Nevertheless, the leading behavior is the universal stretched exponential decay, $\exp(-\text{const.} \times k^{1-\gamma})$; all other terms are subdominant corrections.

Example 14.4. *Superlinear attachment rate.* For $\gamma > 1$, an analog of gelation occurs in which nearly all links condense onto a single node. Ultra-singular behavior occurs for $\gamma > 2$ in which there is a non-zero probability for the existence of a "bible" – a single node that links to *every other* node in an infinite network, while only a finite number of links
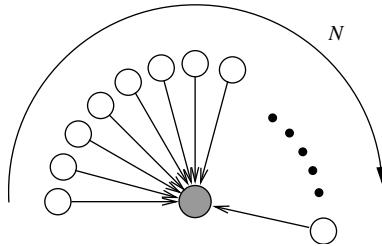
Fig. 14.6. Creation of a "bible" in which each new node attaches only to the bible (shaded).

exist between all other nodes. To determine the probability for a bible to occur, suppose that there is a bible in a network of $N + 1$ nodes (Fig. 14.6). Proceeding inductively, the probability that the next node links to the bible is $N^{\gamma}/(N + N^{\gamma})$, and the probability that this pattern of connections continues indefinitely is $\mathcal{P} = \prod_{N \geq 1}(1 + N^{1-\gamma})^{-1}$. Using the same asymptotic analysis as that used for the sublinear attachment rate, the asymptotic behavior of this product is given by

$$\mathcal{P} = 0, \quad \text{for } \gamma \leq 2,$$
$$\mathcal{P} > 0, \quad \text{for } \gamma > 2.$$

Thus for $\gamma > 2$, there a non-zero probability for a bible to exist in an infinite network.

When $1 < \gamma < 2$, the attachment pattern of low-degree nodes is less trivial than in Fig. 14.6, but there continues to exist a single node that is linked to nearly all nodes. There is also an infinite sequence of transition points when γ passes through $m/(m-1)$, with m an integer greater than two, in which the number of nodes of degree $k < m$ grows as $N^{k-(k-1)\gamma}$, while the number of nodes of larger degree remains finite as the total number of nodes N diverges.

Example 14.5. *Linear attachment rate*. We distinguish between *strictly linear* attachment, $A_k = k$, and *asymptotically linear* attachment. In the former case, the total event rate is $A = \sum_k A_k N_k = \sum_k k N_k = 2N$. Substituting this value of $\mu = 2$ in Eq. (14.36) immediately leads to the discrete power-law form

$$n_k = \frac{4}{k(k+1)(k+2)} = \frac{4\,\Gamma(k)}{\Gamma(k+3)}\,, \tag{14.39}$$

where Γ is the Euler gamma function. Because this distribution has no characteristic degree scale, such networks have been dubbed *scale-free*, and they stand in stark contrast to the delta-function degree distribution of regular lattices and the Poisson degree distribution of the Erdős–Rényi random graph.

For asymptotically linear attachment, the surprising feature is that the degree distribution exponent is *non-universal*. This non-universality is at odds with the common wisdom of statistical physics that the absence of a characteristic scale at a phase transition leads to universal power-law scaling properties. The non-universal behavior for the degree distribution

can be easily derived for asymptotically linear attachment rates, $A_k \sim k$. For this case, Eq. (14.36) becomes

$$n_k = \frac{\mu}{A_k} \prod_{1 \leq j \leq k} \left(1 + \frac{\mu}{A_j}\right)^{-1}$$

$$\sim \frac{\mu}{k} \exp\left[-\int_1^k \ln\left(1 + \frac{\mu}{j}\right) dj\right]$$

$$\sim k^{-(1+\mu)} . \tag{14.40}$$

Thus the degree distribution exponent can take *any* value larger than 2 merely by tuning the amplitude in the total rate $A \simeq \mu N$. Some specific examples of asymptotically linear attachment rates are the subject of problems 14.11–14.15.

14.5 Fluctuations in networks

When we study networks that are generated by a random growth process, we tacitly assume that extensive[12] characteristics of such networks are nearly the same. This assertion cannot be literally true, but it is often morally true. For instance, in an evolving random graph, the number of clusters of a given size is a self-averaging quantity, i.e. its relative fluctuations vanish when the system size diverges (see Eq. (14.12)). It can be hard to verify that a given extensive characteristic is self-averaging, and even harder to establish the nature of the fluctuations even in cases where the characteristic is indeed self-averaging. We now illustrate the possible behaviors of extensive characteristics for two different network growth processes.

Random recursive trees

Let us examine the behavior of the total number $N_1(N)$ of nodes of degree one (such nodes are often called leaves or dangling ends). The average number of leaves is an extensive quantity; from Eq. (14.31) every second node in an RRT is a leaf; that is, $\langle N_1 \rangle \simeq N/2$. A comprehensive description of the random quantity $N_1(N)$ is provided by the probability distribution $P(N_1, N)$. If N_1 is self-averaging, the probability distribution $P(N_1, N)$ will be sharply peaked around $N_1 = N/2$ and very small when N_1 is far from $N/2$. While it is hard to determine $P(N_1, N)$ in the tail regions, the behaviors at the extremes of $N_1 = N - 1$ (star graph) and $N_1 = 2$ (linear chain) are readily computable.

What is the probability $S_N = P(N - 1, N)$ to generate the star graph? For the initial state of a single node, an RRT is certainly a star up to $N = 3$. For larger RRTs, this probability

---

[12] An extensive quantity grows linearly with system size; a standard example is the energy in a system with short-ranged interactions. (Faster-than-linear growth does occur when long-range interactions exist; e.g. the gravitational energy of a system of mass M scales as $M^2/R \sim M^{5/3}$.)

may be found from the recurrence

$$S_{N+1} = \frac{1}{N} S_N,$$

that ensures that the new node links to the center of the star. Thus $S_N = 2/(N-1)!$ for $N \geq 3$. Similarly the probability $\mathcal{L}_N = P(2, N)$ that the RRT process will give a linear chain satisfies the recurrence, for $N \geq 2$,

$$\mathcal{L}_{N+1} = \frac{2}{N} \mathcal{L}_N.$$

Hence $\mathcal{L}_N = 2^{N-2}/(N-1)!$ for $N \geq 3$. Thus for $N \geq 3$, the limiting values of the probability distribution $P(N_1, N)$ are given by

$$P(N_1, N) = \begin{cases} 1/(N-1)!, & \text{star} \ (N_1 = N-1), \\ 2^{N-2}/(N-1)!, & \text{linear chain} \ (N_1 = 2). \end{cases} \tag{14.41}$$

The probabilities of generating these two extreme trees quickly decay as the network size grows.

While we haven't proved anything, the two things we have learned – that the number of leaves N_1 is an extensive random quantity and that the probabilities that N_1 attains extremal values vanish very rapidly with system size – suggest that fluctuations of N_1 are small. A bolder assumption is that fluctuations are Gaussian. Namely, their magnitude scales as $\sqrt{N}$, so that the variance grows linearly with N,

$$\sigma^2(N) = \langle N_1^2(N) \rangle - \langle N_1(N) \rangle^2 \sim N, \tag{14.42a}$$

and the shape is asymptotically Gaussian,

$$P(N_1, N) \simeq \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(N_1 - \langle N_1 \rangle)^2}{2\sigma^2}\right]. \tag{14.42b}$$

This type of behavior is extremely robust and applies well beyond the formal applicability of the central limit theorem.

To test the validity of the Gaussian assumption, let us first try to confirm (14.42a) analytically. Taking the square of (14.25) and averaging we obtain

$$\left\langle N_1^2(N+1) \right\rangle = \left(1 - \frac{2}{N}\right) \langle N_1^2(N) \rangle + \left(2 - \frac{1}{N}\right) \langle N_1(N) \rangle + 1. \tag{14.43}$$

It is easier to work with the variance $\sigma^2(N) = \langle N_1^2(N) \rangle - \langle N_1(N) \rangle^2$. Using Eq. (14.26), we obtain

$$\sigma^2(N+1) = \left(1 - \frac{2}{N}\right) \sigma^2(N) + \frac{1}{N} \langle N_1(N) \rangle - \frac{1}{N^2} \langle N_1(N) \rangle^2. \tag{14.44}$$

Next, we substitute the result for $\langle N_1(N) \rangle$ into (14.44) to recast this equation into a solvable recurrence relation. However, the exact solution to this recurrence is cumbersome, and

because we are primarily interested in asymptotic behavior, let us consider the $N \to \infty$ limit. Then $N^{-1} \langle N_1(N) \rangle \to 1/2$ and we can employ the continuum approximation to reduce the difference equation (14.44) into the differential equation

$$\frac{d\sigma^2}{dN} = -\frac{2\sigma^2}{N} + \frac{1}{4},$$

(14.45)

whose solution is $\sigma^2 = N/12$.

To show that the distribution of fluctuations is Gaussian, we apply the small-fluctuation expansion that was first presented in Section 12.3. This analysis is somewhat tedious but the idea is simple – we merely rewrite the growth rule (14.25) in terms of the probability distribution $P(N_1, N)$,

$$P(N_1, N+1) = \frac{N_1}{N} P(N_1, N) + \left(1 - \frac{N_1 - 1}{N}\right) P(N_1 - 1, N),$$

(14.46)

and then employ the continuum approach, which allows us to use Taylor expansions

$$P(N_1, N+1) = P(N_1, N) + \frac{\partial}{\partial N} P(N_1, N) + \frac{1}{2} \frac{\partial^2}{\partial N^2} P(N_1, N) + \cdots,$$

$$P(N_1 - 1, N) = P(N_1, N) - \frac{\partial}{\partial N_1} P(N_1, N) + \frac{1}{2} \frac{\partial^2}{\partial N_1^2} P(N_1, N) + \cdots.$$

It is now useful to change variables from (N_1, N) to (ν, N) with

$$\nu = N_1 - \frac{N}{2}.$$

(14.47)

In the important region around the peak of the distribution, we have $\nu \sim \sqrt{N}$ and therefore we keep the second derivative $\partial^2/\partial \nu^2$, which is of the order $\partial/\partial N$, but drop $\partial^2/\partial \nu \partial N \sim N^{-3/2}$, $\partial^2/\partial N^2 \sim N^{-2}$, and other higher derivatives. Following these steps, we recast (14.46) into the partial differential equation

$$\frac{\partial \Pi}{\partial N} = \frac{1}{N} \Pi + \frac{\nu}{N} \frac{\partial \Pi}{\partial \nu} + \frac{1}{8} \frac{\partial^2 \Pi}{\partial \nu^2}$$

(14.48)

for $\Pi(\nu, N) = P(\nu + N/2, N)$. General experience with diffusion equations suggests seeking a solution as a function of a single scaling variable $\eta = \nu/\sqrt{N}$, that is

$$\Pi(\nu, N) = \frac{1}{\sqrt{N}} \Phi(\eta),$$

(14.49)

where the $N^{-1/2}$ prefactor ensures the normalization $\int d\nu \, \Pi(\nu, N) = 1$. Substituting (14.49) into (14.48), we obtain the ordinary differential equation whose solution is $\Phi(\eta) = \sqrt{6/\pi} \, e^{-6\eta^2}$ (problem 14.16). We thus have the Gaussian distribution (14.42b) with variance $\sigma^2 = N/12$.

Friendship network

Here we consider a generalization of the RRT where each new node links to a target node and also attempts to link to the neighbors ("friends") of the target.[13] At each time step, a new node is introduced and it connects to a target node that is randomly chosen among already existing nodes. Further, with probability p, the new node also connects to all neighbors of the target.[14] This friendship or copying model therefore depends on the single parameter, the copying probability p. The simplest global network characteristic is the number of links L_N. In contrast to the RRT where L_N is deterministic ($L_N = N - 1$), in the copying model L_N is a random variable. Let's try to compute its average $L(N) = \langle L_N \rangle$ and its variance $\sigma^2(N) = \langle L_N^2 \rangle - [L(N)]^2$. Adding a new node increases on average the total number of links by $1 + p\langle k \rangle$, where $\langle k \rangle = 2L(N)/N$ is the average degree of the network. Therefore the average total number of links evolves as

$$L(N + 1) = L(N) + 1 + 2p \frac{L(N)}{N} \; . \tag{14.50}$$

Within a continuum approach ($N \gg 1$), the difference equation (14.50) becomes the differential equation

$$\frac{dL}{dN} = 1 + 2p \frac{L}{N} \; . \tag{14.51}$$

For the RRT ($p = 0$), we have $L = N - 1$, and we anticipate extensive behavior, $L \sim N$, to hold for sufficiently small p. Substituting this linear dependence into (14.51), we find $L = (1 - 2p)^{-1}N$. Thus extensive behavior is valid when $p < 1/2$. In the marginal case of $p = 1/2$, the solution to (14.51) acquires a logarithmic prefactor, while for $p > 1/2$ the number of links grows faster than linearly:[15]

$$L(N) \simeq \begin{cases} (1 - 2p)^{-1}N, & p < 1/2, \\ N \ln N, & p = 1/2, \\ A(p) N^{2p}, & 1/2 < p \leq 1. \end{cases} \tag{14.52}$$

The important outcome here is that, for $p < \frac{1}{2}$, the mean degree of the network remains finite as $N \to \infty$; equivalently, the number of links is an extensive variable. Such networks are termed *sparse*. Conversely, for $p \geq \frac{1}{2}$, we obtain *dense* networks in which the mean degree diverges as $N \to \infty$.

Let us now compute the variance in the number of links. By construction, a new node attaches to a randomly selected target node and either all or none of its neighbors.

---

[13] Copying is a key mechanism in the growth of the World-Wide Web, the citation network, and other social, technological, and information networks. Copying also occurs in Nature; e.g. gene duplication, which plays a crucial role in evolution, is essentially based on copying.

[14] A variant that exhibits qualitatively similar behavior is one where attempts to connect to neighbors are independent and each is successful with probability p.

[15] The continuum approach gives the correct dependence on N and the correct amplitude for $p \leq 1/2$. For $p > 1/2$, the continuum approach gives only the exponent, $L \sim N^{2p}$, and one has to solve the exact recurrence (14.50) to determine the amplitude $A(p)$.

That is

$$L_{N+1} = L_N + 1 + \begin{cases} k, & \text{probability } p, \\ 0, & \text{probability } 1-p, \end{cases} \tag{14.53}$$

where k is the degree of the target node. Taking the square and averaging, we obtain

$$\langle L_{N+1}^2 \rangle = p\langle (L_N + 1 + k)^2 \rangle + (1-p)\langle (L_N + 1)^2 \rangle$$
$$= \langle (L_N + 1)^2 \rangle + 2p\langle L_N k \rangle + 2p\langle k \rangle + p\langle k^2 \rangle. \tag{14.54}$$

We know that $\langle k \rangle = 2L(N)/N$, while $\langle L_N k \rangle = \langle L_N \sum k \rangle / N$. However, $\langle \sum k \rangle = 2L_N$, so that we obtain $\langle L_n k \rangle = 2\langle L_N^2 \rangle / N$. Thus (14.54) becomes

$$\langle L_{N+1}^2 \rangle = \left(1 + \frac{4p}{N}\right)\langle L_N^2 \rangle + 2\left(1 + \frac{2p}{N}\right)L(N) + 1 + p\langle k^2 \rangle.$$

Subtracting the square of (14.50) from this equation gives

$$\sigma^2(N+1) = \left(1 + \frac{4p}{N}\right)\sigma^2(N) + p\langle k^2 \rangle - p^2\langle k \rangle^2. \tag{14.55}$$

In the continuum approximation Eq. (14.55) turns into the differential equation

$$\frac{d\sigma^2}{dN} = \frac{4p}{N}\sigma^2 + p\langle k^2 \rangle - p^2\langle k \rangle^2$$

whose solution is

$$\sigma^2 \simeq \frac{p\langle k^2 \rangle - p^2\langle k \rangle^2}{1 - 4p}N, \qquad p < 1/4. \tag{14.56}$$

When $1/4 \le p \le 1$, we find faster-than-linear growth[16]

$$\sigma^2 \sim N^{4p}. \tag{14.57}$$

For $p > 1/2$, the average number of links and its standard deviation scale in the same way, $L \sim \sigma \sim N^{2p}$. Therefore dense networks that are built by copying strongly fluctuate from realization to realization.

Example 14.6. *Urn model.* Consider the copying model with $p = 1$, i.e. the new node chooses a target node and attaches to it and to all its neighbors. Suppose that the network initially contains two disconnected nodes.[17] The emergent network will always consist of two disjoint complete graphs, say of size A and B with $A + B = N$. What is the fraction of nodes $A/(A+B)$ in one of these graphs? Remarkably, in each realization, this fraction approaches a limiting random value that is *uniformly* distributed between zero and one.[18]

---

[16] Plus logarithmic corrections at $p = 1/4$ and $p = 1/2$.

[17] If the initial network consists of a single node, the emergent network is the complete graph.

[18] This model is known as the Eggenberger–Pólya urn model: one starts with two marbles of different colors in an urn, draws a marble randomly from the urn, and puts it back, together with another marble of the same color.

To understand this outcome, let's denote by θ_N the fraction of nodes in one subgraph and examine its evolution as a function of the network size N. Initially $\theta_2 = 1/2$. At the next step, $\theta_3 = 1/3$ or $\theta_3 = 2/3$ with equal probabilities. Then we get $\text{Prob}(\theta_4 = 1/4) = \text{Prob}(\theta_4 = 2/4) = \text{Prob}(\theta_4 = 3/4) = 1/3$. Generally $\text{Prob}(\theta_N = j/N) = (N-1)^{-1}$ for all $j = 1, \ldots, N-1$. In a given realization, the fraction θ_N fluctuates less and less as N increases and eventually it saturates at a certain limiting value. The limiting value, however, is totally unpredictable; namely, it is uniformly distributed on the interval $(0, 1)$. The very first steps of the growth are particularly important as they lock in initial fluctuations. This is a striking example of the lack of self-averaging.

14.6 Notes

The network reproduced in Fig. 14.1 appear in [261, 262]. Graph theory and complex networks of various types have long been studied by mathematicians. The literature devoted to planar regular random graphs is immense as they are connected with many branches of mathematics and physics, such as combinatorics, random matrices, and toy models in string theory. Good starting points are the review article [263] and the book [264]. The paper by Erdős and Rényi [265] on random graphs provides the mathematical motivation for this surprisingly rich subject. The most spectacular feature of the random graph is the emergence of the giant component as the density of links increases beyond a critical value. A useful analysis of this phenomenon is given in the review article [266]. Two useful texts on random graphs are [267, 268]. The construction of the scaling window was developed in [269]. Various geometrical properties of random graphs from a kinetic viewpoint are discussed in [270].

An early work about basic geometrical properties of random recursive trees is given in [271], and a review of this subject is given in [272]. The seeds of preferential attachment networks appears in work by Yule on species abundance distributions [273]. Perhaps the earliest work on preferential attachment networks *per se* is that by Simon [274]. This class of networks gained considerable popularity after it was found that such networks mimic many properties of technological and real-life biological networks. A review of this work is given in [275] while recent monographs on this subject include [276–278]. The master equation approach to solve for the degree distribution and related geometrical properties was developed in [279]. For more on the influence of finiteness on the structure of networks, see [280, 281].

14.7 Problems

14.1 Show that regular random graphs with $z = 1$ and $z = 2$ can be defined dynamically and the planarity condition is automatically obeyed. For $z = 1$, the graph is a collection of dimers, while for $z = 2$, the graph is a collection of loops. Show that

for $z = 1$ the final number of components is $N/2$, while for $z = 2$ the final number of components scales logarithmically with system size.

14.2 Solve the evolution equations (14.1) for the degree distribution of the ER graph, subject to the initial condition $n_k(t = 0) = \delta_{k,0}$, and show that the solution is given by (14.2). Try to do it by different methods (directly solving the equations one by one, the Laplace transform method, and the generating function technique).

14.3 Derive the governing equations for the moments $U_1 = \sum_{k \geq 1} k u_k$ and $U_2 = \sum_{k \geq 1} k^2 u_k$ and solve these equations. Show that in the sol phase

$$U_1 = \frac{t}{2(1-t)^2}, \quad U_2 = \frac{t(1+t)}{2(1-t)^4}.$$

14.4 Find a solution to the recurrence (14.18). One way is via generating functions:

(a) Using the exponential generating functions, $\mathcal{A}(z) = \sum_{k \geq 1} A_k e^{kz}$ and $\mathcal{H}(z) = \sum_{k \geq 1} (k^{k-1}/k!) e^{kz}$, recast the recurrence (14.18) into the differential equation

$$\mathcal{A} = (1 - \mathcal{H})^{-1} \frac{d\mathcal{H}}{dz}.$$

(b) The generating function $\mathcal{H}(z)$ encapsulates the cluster size distribution at the percolation point. Using the results from Chapter 5 show that $\mathcal{H} e^{-\mathcal{H}} = e^z$. Combine this with the above differential equation to yield $\mathcal{A} = (1 - \mathcal{H})^{-2} \mathcal{H}$.

(c) Employ the Lagrange inversion method to compute A_k.

14.5 Determine the large-k asymptotic behavior of the sum $\sum_{0 \leq n \leq k-1} k^{n-1}/n!$. Here is one possible line of attack:

(a) Notice that the last term in the sum is the largest. This suggests reversing the summation order and writing the sum in the form

$$\frac{k^{k-2}}{(k-1)!} S, \quad S = 1 + \frac{k-1}{k} + \frac{(k-1)(k-1)}{k^2} + \frac{(k-1)(k-1)(k-3)}{k^3} + \cdots$$

(b) To compute S show that in the large-k limit it is possible to use the continuum approximation to give

$$S = \sum_{j=0}^{k-1} \prod_{i=0}^{j} \left(1 - \frac{i}{k}\right) \rightarrow \int_0^k dj \, \exp\left[-\frac{j(j+1)}{2k}\right]$$

$$\rightarrow \int_0^\infty dj \, \exp\left[-\frac{j^2}{2k}\right] = \sqrt{\pi k/2}.$$

(c) Combining (a) and (b) and using Stirling's formula $k! \simeq (k/e)^k \sqrt{2\pi k}$ show that

$$\sum_{0 \leq n \leq k-1} \frac{k^{n-1}}{n!} = \frac{k^{k-1}}{k!} S \simeq \frac{k^k}{k!} \sqrt{\frac{\pi}{2k}} \simeq \frac{e^k}{2k}$$

and from this result deduce (14.21).

14.6 Show that, for a large but finite evolving random graph, the probability $s_n(N)$ to have n unicyclic clusters at the percolation point is given by

$$s_n(N) \sim \frac{1}{n!} N^{-1/6} \left[\frac{1}{6} \ln N \right]^n .$$

14.7 Compute the exact average degree distribution $\langle N_k(N) \rangle$ for the random recursive tree for the first few $N = 1, 2, 3, \ldots$, by solving the recursion

$$\langle N_k(N+1) \rangle - \langle N_k(N) \rangle = \frac{\langle N_{k-1}(N) \rangle - \langle N_k(N) \rangle}{N} + \delta_{k,1}.$$

Here you should apply the generating function $G_k(w) = \sum_{N \geq 1} \langle N_k(N) \rangle w^{N-1}$ to convert this recursion into soluble equations. Next, expand the solutions for $G_k(w)$ in power series in N to obtain the degree distributions. In particular, show that the average number of nodes of degree one is given by (14.27). Finally, compare your results with the asymptotic average degree distribution $N_k(N) \simeq N/2^k$.

14.8 Extend the genealogical tree construction to the RRT with redirection.

(a) First show that the master equation for L_g, the number of nodes in generation g, is given by

$$\frac{dL_g}{dN} = \frac{(1-r)L_{g-1} + rL_g}{N} .$$

(b) Show that the solution for L_g is

$$L_{g+1}(\tau) = \int_0^\tau dx \, \frac{[(1-r)x]^g}{g!} e^{xr} ,$$

where $\tau = \ln N$.

14.9 Show that the limiting behavior of μ in $A(N) = \mu N$ is given by

$$\mu = 1 + B_0 \gamma + \mathcal{O}\left(\gamma^2\right), \qquad\qquad \gamma \downarrow 0,$$

$$\mu = 2 - B_1(1-\gamma) + \mathcal{O}\left((1-\gamma)^2\right), \qquad \gamma \uparrow 1,$$

with

$$B_0 = \sum_{j \geq 1} \frac{\ln j}{2^j} \doteq 0.5078, \qquad B_1 = 4 \sum_{j \geq 1} \frac{\ln j}{(j+1)(j+2)} \doteq 2.407.$$

Here γ is the exponent in the attachment rate A_k defined by $A_k = k^\gamma$.

14.10 For linear preferential attachment, $A_k = k$, calculate N_k explicitly from the master equations (14.34) for $k = 1, 2, 3$, and 4.

14.11 Determine the degree distribution for the shifted linear attachment rate $A_k = k + \lambda$. First simplify the normalization factor $A = \sum_j A_j N_j$ to $A(N) = M_1 + \lambda M_0 =$

$(2 + \lambda)N$, implying that $\mu = 2 + \lambda$. Using these results show that the degree distribution is

$$n_k = (2 + \lambda) \frac{\Gamma(3 + 2\lambda)}{\Gamma(1 + \lambda)} \frac{\Gamma(k + \lambda)}{\Gamma(k + 3 + 2\lambda)}.$$

Show that asymptotically this distribution decays as $k^{-3-\lambda}$. Notice also that Eq. (14.34) with the shifted linear attachment rate $A_k = k + \lambda$ is *identical* to that of the redirection model, Eq. (14.32), when the correspondence $\lambda = 1/r - 2$ is made.

14.12 Determine the degree distribution for the growth process in which, with probability p, a new node attaches to an existing node of the network by linear preferential attachment and, with probability $1 - p$, a new node attaches to an existing node by uniform random attachment.

14.13 Consider the connection kernel $A_1 = 1$ and $A_k = ak$ for $k \geq 2$. Show that the resulting degree distribution is asymptotically a power law, $n_k \sim k^{-\nu}$, with $\nu = (3 + \sqrt{1 + 8/a})/2$, which can indeed be tuned to any value larger than 2.

14.14 Generalize linear preferential attachment networks to the case where each new node links to m pre-existing nodes. Write the master equation for this process, and by applying the same approach as that used for Eq. (14.34), find the degree distribution.

14.15 Suppose that each node is assigned an initial "attractiveness" $\eta > 0$ from a distribution $p_0(\eta)$. The attachment rate of a node with degree k and attractiveness η is defined as $A_k(\eta)$. Let $N_k(\eta)$ be the number of nodes of degree k and attractiveness η. Show that this joint degree–attractiveness distribution evolves according to

$$\frac{dN_k(\eta)}{dN} = \frac{A_{k-1}(\eta)N_{k-1}(\eta) - A_k(\eta)N_k(\eta)}{A} + p_0(\eta)\delta_{k1},$$

where $A = \int d\eta \sum_k A_k(\eta)N_k(\eta)$ is the total attachment rate. By using the substitutions $A = \mu N$ and $N_k(\eta) = N n_k(\eta)$ reduce the above master equation to a recursion whose solution is

$$n_k(\eta) = p_0(\eta)\frac{\mu}{A_k(\eta)} \prod_{1 \leq j \leq k} \left(1 + \frac{\mu}{A_j(\eta)}\right)^{-1}.$$

For the specific example where the attachment rate is linear in the degree *and* attractiveness, $A_k(\eta) = \eta k$, show that the degree distribution is

$$n_k(\eta) = \frac{\mu \, p_0(\eta)}{\eta} \frac{\Gamma(k) \, \Gamma(1 + \mu/\eta)}{\Gamma(k + 1 + \mu/\eta)}.$$

14.16 Fill the gaps in the derivation of the Gaussian distribution that was outlined in (14.46)–(14.49). In particular, derive the differential equation for the scaling function $\Phi(\eta)$ and solve it.

14.17 The goal of this problem is to solve Eq. (14.50).

(a) Show that the homogeneous version of (14.50) admits the following solution:

$$\prod_{1 \le j \le N} \left(1 + \frac{2p}{j}\right) = \frac{\Gamma(2p + N)}{\Gamma(2p + 1)\,\Gamma(N)}.$$

(b) Argue that it is a good idea to use the solution of the homogeneous version of (14.50) as an integrating factor, that is, to seek the solution in the form

$$L(N) = U(N)\,\frac{\Gamma(2p + N)}{\Gamma(2p + 1)\,\Gamma(N)}.$$

Show that the above ansatz recasts (14.50) into

$$U(N + 1) = U(N) + \frac{\Gamma(2p + 1)\,\Gamma(N + 1)}{\Gamma(2p + N + 1)}.$$

(c) Solve the above recurrence to yield

$$L(N) = \frac{\Gamma(2p + N)}{\Gamma(N)} \sum_{2 \le j \le N} \frac{\Gamma(j)}{\Gamma(2p + j)}.$$

(d) Using the well-known asymptotic relation

$$\frac{\Gamma(2p + x)}{\Gamma(x)} \to x^{2p} \quad \text{when} \quad x \gg 1,$$

confirm the asymptotic behavior (14.52) when $p < 1/2$.

(e) Show that, when $p = 1/2$, the exact result is $L(N) = N(H_N - 1)$ where $H_N = \sum_{1 \le j \le N} j^{-1}$ is the harmonic number. Using the asymptotics of the harmonic numbers confirm the prediction of Eq. (14.52) for $p = 1/2$.

(f) Show that the sum on the right-hand side of the equation in (c) converges. Compute this sum using the identity

$$\sum_{k \ge 0} \frac{\Gamma(b + k)}{\Gamma(c + k)} = \frac{\Gamma(b)}{(c - b - 1)\,\Gamma(c - 1)}.$$

Confirm Eq. (14.52) for $p > 1/2$ with $A(p) = 1/[(2p - 1)\,\Gamma(1 + 2p)]$.

References

Chapter 1: Aperitifs

[1] W. Feller, *An Introduction to Probability Theory and its Applications*, Vol. I (Wiley, New York, 1968).

[2] S. Karlin and H. M. Taylor, *A First Course in Stochastic Processes*, 2nd edn. (Academic Press, New York, 1975).

[3] G. H. Weiss, *Aspects and Applications of the Random Walk* (North-Holland, Amsterdam, 1994).

[4] J. Rudnick and G. Gaspari, *Elements of the Random Walk* (Cambridge University Press, New York, 2004).

[5] P. Mörders and Y. Peres, *Brownian Motion*, http://www.stat.berkeley.edu/users/peres/bmbook.pdf.

[6] B. Duplantier, in: *Einstein, 1905–2005*, eds Th. Damour, O. Darrigol, B. Duplantier and V. Rivasseau (Birkhäuser, Basel, 2006); arXiv:0705.1951.

[7] Galileo Galilei, *Discourses and Mathematical Demonstrations Relating to Two New Sciences* (Dover, New York, 1954; originally published by Elzevir, Leiden, 1638).

[8] Yu. I. Manin, *Mathematics and Physics* (Birkhäuser, Basel, 1981).

[9] G. I. Barenblatt, *Scaling, Self-Similarity, and Intermediate Asymptotics* (Cambridge University Press, Cambridge, 1996).

[10] N. Goldenfeld, *Lectures on Phase Transitions and the Renormalization Group* (Addison-Wesley, Reading, MA, 1992).

[11] P. G. de Gennes, *Scaling Concepts in Polymer Physics* (Cornell University Press, Ithaca, NY, 1979).

Chapter 2: Diffusion

[12] S. Chandrasekhar, *Rev. Mod. Phys*. **15**, 1 (1943).

[13] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 2001).

[14] S. Alexander, J. Bernasconi, W. R. Schneider, and R. Orbach, *Rev. Mod. Phys.* **53**, 175 (1981).

[15] J.-P. Bouchaud and A. Georges, *Phys. Rep.* **195**, 127 (1990).

[16] B. Derrida, *Physica D* **107**, 186 (1997).

[17] S. Redner, *A Guide to First-Passage Processes* (Cambridge University Press, Cambridge, 2001).

[18] M. E. Fisher and M. P. Gelfand, *J. Stat. Phys.* **53**, 175 (1988).

[19] M. Bramson and D. Griffeath, in: *Random Walks, Brownian Motion, and Interacting Particle Systems: A Festshrift in Honor of Frank Spitzer*, eds R. Durrett and H. Kesten (Birkhäuser, Boston, 1991).

[20] D. ben-Avraham, B. M. Johnson, C. A. Monaco, P. L. Krapivsky, and S. Redner, *J. Phys. A* **36**, 1789 (2003).

[21] H. C. Berg and E. M. Purcell, *Biophys. J.* **20**, 193 (1977).

[22] H. C. Berg, *Random Walks in Biology* (Princeton University Press, Princeton, NJ, 1993).

[23] H. S. Carslaw and J. C. Jaeger, *Conduction of Heat in Solids* (Clarendon Press, Oxford, 1959).

[24] J. Crank, *Free and Moving Boundary Problems* (Oxford University Press, Oxford, 1984).

[25] J. S. Langer, in: *Chance and Matter*, eds J. Souletie, J. Vannimenus, and R. Stora (North-Holland, Amsterdam, 1987).

[26] S. Redner and D. ben-Avraham, *J. Phys. A* **23**, L1169 (1990).

[27] P. L. Krapivsky, *Phys. Rev. E* **47**, 1199 (1993).

[28] T. Halpin-Healey and Y.-C. Zhang, *Phys. Rep.* **254**, 215 (1995).

[29] A.-L. Barabasi and H. E. Stanley, *Fractal Concepts in Surface Growth* (Cambridge University Press, Cambridge, 1995).

[30] M. Kardar, in: Les Houches session LXII on *Fluctuating Geometries in Statistical Mechanics and Field Theory*, eds F. David, P. Ginzparg, and J. Zinn-Justin (Elsevier, Amsterdam, 1996); arXiv:cond-mat/9411022.

Chapter 3: Collisions

[31] J. C. Maxwell, *Phil. Trans. Roy. Soc.* **157**, 49 (1867).

[32] L. Boltzmann, *Lectures on Gas Theory* (Dover, New Yok, 1964).

[33] M. Kac, *Probability and Related Topics in Physical Sciences* (Interscience, London, 1959).

[34] P. Resibois and M. De Leener, *Classical Kinetic Theory of Fluids* (Wiley, New York, 1977).

[35] C. Truesdell and R. G. Muncaster, *Fundamentals of Maxwell's Kinetic Theory of a Simple Monoatomic Gas* (Academic Press, New York, 1980).

[36] M. H. Ernst, *Phys. Rep.* **78**, 1 (1981).

[37] C. Villani, A review of mathematical topics in collisional kinetic theory in: *Handbook of Mathematical Fluid Dynamics*, eds S. Friedlander and D. Serre (Elsevier Science, Amsterdam, 2002).

[38] K. Huang, *Statistical Mechanics* (Wiley, New York, 1987).

[39] P. Drude, *Annalen Phys*. **1**, 566 (1900); **3**, 369 (1900).

[40] H. A. Lorentz, *Arch. Neerl.* **10**, 336 (1905); reprinted in: *Collected Papers* Vol. III, p. 180 (Martinus Nijhoff, The Hague, 1936).

[41] Sometimes, the Lorentz gas is called a Galton board, see *e.g.* M. Kac, *Scientific American* **211**, 2, 92 (1964).

[42] E. H. Hauge, in: *Transport Phenomena*, eds G. Kirczenow and J. Marro, Lecture Notes in Physics, Vol. 31, p. 337 (Springer, Berlin, 1974).

[43] L. A. Bunimovich and Ya. G. Sinai, *Comm. Math. Phys.* **78**, 247 (1980); L. A. Bunimovich and Ya. G. Sinai, *Comm. Math. Phys.* **78**, 479 (1981).

[44] J. Piasecki and E. Wajnryb, *J. Stat. Phys.* **21**, 549 (1979).

[45] R. Kubo, M. Toda, and N. Hashitsumi, *Nonequilibrium Statistical Mechanics* (Springer, Berlin, 2003).

[46] T. Antal, P. L. Krapivsky, and S. Redner, *Phys. Rev. E* **78**, 030301 (2008); (arXiv:0805.3783).

[47] L. I. Sedov, *Similarity and Dimensional Methods in Mechanics* (Academic Press, New York, 1959).

[48] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics* (Pergamon Press, New York, 1987).

[49] P. K. Haff, *J. Fluid Mech.* **134**, 401 (1983).

[50] I. Goldhirsch and G. Zanetti, *Phys. Rev. Lett.* **73**, 1619 (1993).

[51] A. Baldassarri, U. M. B. Marconi, and A. Puglisi, *Europhys. Lett.* **58**, 14 (2002).

[52] N. Brilliantov and T. Pöschel, *Kinetic Theory of Granular Gases* (Oxford University Press, Oxford, 2004).

[53] R. S. Krupp, *A nonequilibrium solution of the Fourier transformed Boltzmann equation*, M.S. Thesis, MIT (1967); *Investigation of solutions to the Fourier transformed Boltzmann equation*, Ph.D. Thesis, MIT (1970).

[54] A. V. Bobylev, *Sov. Sci. Rev. C. Math. Phys.* **7**, 111 (1988).

[55] E. Ben-Naim and P. L. Krapivsky, The inelastic Maxwell model, in: *Granular Gases*, Lecture Notes in Physics, Vol. 624, p. 63 (Springer, Berlin, 2004); arXiv:cond-mat/0301238.

[56] J. M. Burgers, *The Nonlinear Diffusion Equation* (Reidel, Dordrecht, 1974).

[57] S. Kida, *J. Fluid Mech.* **93**, 337 (1979).

[58] S. F. Shandarin and Y. B. Zeldovich, *Rev. Mod. Phys.* **61**, 185 (1989).

[59] G. F. Carnevale, Y. Pomeau, and W. R. Young, *Phys. Rev. Lett.* **64**, 2913 (1990).

[60] E. Trizac and P. L. Krapivsky, *Phys. Rev. Lett.* **91**, 218302 (2003); arXiv:cond-mat/0310133.

[61] L. Frachebourg, *Phys. Rev. Lett.* **82**, 1502 (1999).

[62] D. C. Rapaport, *The Art of Molecular Dynamics Simulation*, 2nd edn. (Cambridge University Press, Cambridge, 2004).

[63] W. H. Press, S. A. Teukolsky, W. T. Vetterlin, and B. P. Flannery, *Numerical Recipes*, 2nd edn. (Cambridge University Press, Cambridge, 1992).

[64] E. Ben-Naim, P. L. Krapivsky, and S. Redner, *Phys. Rev. E* **50**, 822 (1994).

Chapter 4: Exclusion

[65] A. L. Hodgkin and R. D. Keynes, *J. Phys.* **128**, 61 (1955).

[66] T. E. Harris, *J. Appl. Prob.* **2**, 323 (1965).

[67] F. Spitzer, *Adv. Math.* **5**, 246 (1970).

[68] D. G. Levitt, *Phys. Rev. A* **8**, 3050 (1973).

[69] G. B. Whitham, *Linear and Nonlinear Waves* (Wiley, New York, 1974).

[70] D. J. Logan, *An Introduction to Nonlinear Partial Differential Equations* (Wiley, New York, 1994).

[71] T. M. Liggett, *Stochastic Interacting Systems: Contact, Voter, and Exclusion Processes* (Springer, New York, 1999).

[72] B. Schmittmann and R. K. P. Zia, Statistical mechanics of driven diffusive systems, in: *Phase Transitions and Critical Phenomena*, Vol. 17, eds C. Domb and J. L. Lebowitz (Academic Press, London, 1995).

[73] *Nonequilibrium Statistical Mechanics in One Dimension*, ed. V. Privman (Cambridge University Press, New York, 1997). This book contains many useful reviews. For exclusion, see B. Derrida and M. R. Evans, The asymmetric exclusion model: exact results through a matrix approach, chap. 14.

[74] G. Schütz, Exactly solvable models for many-body systems far from equilibrium, in: *Phase Transitions and Critical Phenomena*, Vol. 19, eds C. Domb and J. L. Lebowitz (Academic Press, London, 2000).

[75] O. Golinelli and K. Mallick, *J. Phys. A* **39**, 12679 (2006); arXiv:cond-mat/0611701.

[76] R. A. Blythe and M. R. Evans, *J. Phys. A* **40**, R333 (2007); arXiv:0706.1678.

[77] B. Derrida, *Phys. Rep.* **301**, 65 (1998).

[78] B. Derrida, *J. Stat. Mech.* P07023 (2007); arXiv:cond-mat/0703762.

[79] M. Lifshits, S. A. Gredeskul, and L. A. Pastur, *Introduction to the Theory of Disordered Systems* (Wiley, New York, 1988).

Chapter 5: Aggregation

[80] P. J. Flory, *Principles of Polymer Chemistry* (Cornell University Press, Ithaca, NY, 1953).

[81] R. L. Drake, in: *Topics in Current Aerosol Researches*, eds G. M. Hidy and J. R. Brock, p. 201 (Pergamon Press, New York, 1972).

[82] S. K. Frielander, *Smoke, Dust and Haze: Fundamentals of Aerosol Behavior* (Wiley, New York, 1977).

[83] H. Pruppacher and J. Klett, *Microphysics of Clouds and Precipitations* (Kluwer, Dordrecht, 1998).

[84] R. M. Ziff, E. D. McGrady, and P. Meakin, *J. Chem. Phys.* **82**, 5269 (1985).

[85] R. Probstein, *Physicochemical Hydrodynamics: An Introduction* (Wiley-Interscience, New York, 2003).

[86] S. K. Lando, *Lectures on Generating Functions*, Student Mathematical Library, Vol. 23 (American Mathematical Society, Providence, RI, 2003).

[87] R. L. Graham, D. E. Knuth, and O. Patashnik, *Concrete Mathematics: A Foundation for Computer Science* (Addison-Wesley, Reading, MA, 1989).

[88] H. S. Wilf, *Generating Functionology* (Academic Press, Boston, 1990).

[89] P. Flajolet and R. Sedgewick, *Analytic Combinatorics* (Cambridge University Press, New York, 2009).

[90] D. J. Aldous, *Bernoulli* **5**, 3 (1999).

[91] F. Leyvraz, *Phys. Rep.* **383**, 95 (2003); arXiv:cond-mat/0305670.

[92] G. B. Field and W. C. Saslaw, *Astrophys. J.* **142**, 568 (1965).

[93] W. H. White, *J. Colloid Interface Sci.* **87**, 204 (1982).

[94] N. V. Brilliantov and P. L. Krapivsky, *J. Phys. A* **24**, 4787 (1991).

[95] L. H. Tang, *J. Phys. I (France)* **3**, 935 (1993).

[96] P. L. Krapivsky, J. F. F. Mendes, and S. Redner, *Phys. Rev. B* **59**, 15950 (1999); arXiv:cond-mat/9808054.

[97] M. N. Popescu, J. G. Amar, and F. Family, *Phys. Rev. B* **64**, 205404 (2001).

[98] Z. A. Melzak, *Mathematical Ideas, Modeling and Applications* (Wiley, New York, 1976).

[99] S. Ispolatov, P. L. Krapivsky, and S. Redner, *Eur. Phys. J. B* **2**, 267 (1998).

[100] E. R. Domilovskii, A. A. Lushnikov, and V. N. Piskunov, *Dokl. Phys. Chem.* **240**, 108 (1978).

[101] P. G. J. van Dongen, *J. Phys. A* **20**, 1889 (1987).

[102] L. Malyushkin and J. Goodman, *Icarus* **150**, 314 (2001).

[103] E. Ben-Naim and P. L. Krapivsky, *Phys. Rev. E* **68**, 031104 (2003); arXiv:cond-mat/0305154.

[104] http://www.research.att.com/∼njas/sequences/

Chapter 6: Fragmentation

[105] See the articles in *Statistical Models for the Fracture of Disordered Media*, eds H. J. Herrmann and S. Roux (Plenum, New York, 1990).

[106] T. Harris, *The Theory of Branching Processes* (Springer, Berlin, 1963).

[107] R. M. Ziff, *J. Stat. Phys.* **23**, 241 (1980).

[108] E. D. McGrady and R. M. Ziff, *Phys. Rev. Lett.* **58**, 892 (1987).

[109] R. M. Ziff and E. D. McGrady, *J. Phys. A* **18**, 3027 (1985).

[110] M. H. Ernst and G. Szamel, *J. Phys. A* **26**, 6085 (1993).

[111] Z. Cheng and S. Redner, *J. Phys. A* **23**, 1233 (1990).

[112] A. F. Filippov, *Theory Prob. Appl.* **6**, 275 (1961).

[113] E. Ben-Naim and P. L. Krapivsky, *Phys. Lett. A* **293**, 48 (2000).

[114] A. Charlesby, *Proc. Roy. Soc. London Ser. A* **224**, 120 (1954).

[115] E. Ben-Naim and P. L. Krapivsky, *Phys. Rev. E* **50**, 3502 (1994).

[116] P. J. Blatz and A. V. Tobolsky, *J. Phys. Chem.* **49**, 77 (1945).

[117] E. Ben-Naim and P. L. Krapivsky, *Phys. Rev. E* **77**, 061132 (2008); arXiv:0803.2875.

[118] P. L. Krapivsky and S. Redner, *Phys. Rev. E* **54**, 3553 (1996); arXiv:cond-mat/9509129.

[119] S. N. Majumdar, S. Krishnamurthy, and M. Barma, *Phys. Rev. Lett.* **81**, 3691 (1998).

[120] P. L. Krapivsky and E. Ben-Naim, *Phys. Rev. E* **68**, 021102 (2003); arXiv:cond-mat/0302525.

Chapter 7: Adsorption

[121] P. J. Flory, *J. Amer. Chem. Soc.* **61**, 1518 (1939).

[122] A. Rényi, *Publ. Math. Inst. Hung. Acad. Sci.* **3**, 109 (1958).

[123] J. J. Gonzalez, P. C. Hemmer, and J. C. Hoye, *Chem. Phys.* **3**, 228 (1974).

[124] M. C. Bartelt and V. Privman, *Int. J. Mod. Phys.* **5**, 2883 (1991).

[125] J. W. Evans, *Rev. Mod. Phys.* **65**, 1281 (1993); see also Random and cooperative sequential adsorption: exactly solvable models on 1D lattices, continuum limits, and 2D extensions, chap. 10 in [73].

[126] J. Talbot, G. Tarjus, P. R. Van Tassel, and P. Viot, *Colloids Surfaces A* **165**, 287 (2000).

[127] C. Monthus and H. J. Hilhorst, *Physica A* **175** 263 (1991).

[128] Y. Pomeau, *J. Phys. A* **13** L193 (1980).

[129] R. H. Swendsen, *Phys. Rev. A* **24**, 504 (1981).

[130] J. Talbot, G. Tarjus, and P. Schaaf, *Phys. Rev. A* **40**, 4808 (1989).

[131] G. Tarjus and P. Viot, *Phys. Rev. Lett.* **67**, 1875 (1991).

[132] P. L. Krapivsky and E. Ben-Naim, *J. Chem. Phys.* **100**, 6778 (1994); X. Jin, G. Tarjus, and J. Talbot, *J. Phys. A* **27**, L195 (1994).

[133] E. R. Nowak, J. B. Knight, E. Ben-Naim, H. Jaeger, and S. Nagel, *Phys. Rev. E* **57**, 1971 (1998).

[134] P. Nielaba, Lattice models of irreversible adsorption and diffusion, chapt. 11 in [73].

[135] S. M. Simon, C. S. Peskin, and G. F. Oster, *Proc. Natl. Acad. Sci. (USA)* **89**, 3770 (1992).

[136] M. R. D'Orsogna, T. Chou, and T. Antal, *J. Phys. A* **40**, 5575 (2007).

[137] G. N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge University Press, Cambridge, 1922).

[138] C. M. Bender and S. A. Orszag, *Advanced Mathematical Methods for Scientists and Engineers* (McGraw-Hill, New York, 1978).

Chapter 8: Spin dynamics

[139] J. D. Gunton, M. San Miguel, and P. S. Sahni, in: *Phase Transitions and Critical Phenomena*, Vol. 8, eds. C. Domb and J. L. Lebowitz (Academic Press, New York, 1983).

[140] A. J. Bray, *Adv. Phys.* **43**, 357 (1994).

[141] T. M. Liggett, *Interacting Particle Systems* (Springer, New York, 1985).

[142] I. Dornic, H. Chaté, J. Chave, and H. Hinrichsen, *Phys. Rev. Lett.* **87**, 045701 (2001).

[143] P. L. Krapivsky, *Phys. Rev. A* **45**, 1067 (1992).

[144] L. Frachebourg and P. L. Krapivsky, *Phys. Rev. E* **53**, R3009 (1996); arXiv:cond-mat/9508123.

[145] R. J. Glauber, *J. Math. Phys.* **4**, 294 (1963).

[146] B. U. Felderhof, *Rep. Math. Phys.* **1**, 215 (1970).

[147] D. Bedeaux, K. E. Shuler, and I. Oppenheim, *J. Stat. Phys.* **2**, 1 (1970).

[148] P. L. Krapivsky and E. Ben-Naim, *Phys. Rev. E* **56**, 3788 (1997); arXiv:cond-mat/9705155.

[149] B. Derrida and R. Zeitek, Phys. Rev. E **54**, 2513 (1996); arXiv:cond-mat/9606042.

[150] S. N. Majumdar, D. S. Dean, and P. Grassberger, *Phys. Rev. Lett.* **86**, 2301 (2001); arXiv:cond-mat/0004486.

[151] V. Spirin, P. L. Krapivsky, and S. Redner, *Phys. Rev. E* **63**, 036118 (2001); *Phys. Rev. E* **65**, 016119 (2001).

[152] K. Barros, P. L. Krapivsky, and S. Redner, *Phys. Rev. E* **80**, 040101(R) (2009); arXiv:0905.3521.

[153] H. Rost, *Theor. Prob. Rel. Fields* **58**, 41 (1981).

[154] L. Chayes, R. H. Schonmann, and G. Swindle, *J. Stat. Phys.* **79**, 821 (1995).

[155] A. Karma and A. E. Lobkovsky, *Phys. Rev. E* **71**, 036114 (2005); arXiv:cond-mat/0502219.

[156] R. Cerf and S. Louhichi, *Prob. Theory Rel. Fields* **137**, 379 (2007); arXiv:math/0411545.

[157] K. Kawasaki, *Phys. Rev.* **145**, 224 (1966).

[158] S. J. Cornell, K. Kaski, and R. B. Stinchcombe, *Phys. Rev. B* **44**, 12263 (1991).

[159] R. H. Swendsen and J. S. Wang, *Phys. Rev. Lett.* **58**, 86 (1987).

[160] U. Wolff, *Phys. Rev. Lett.* **62**, 361 (1989).

[161] D. P. Landau and K. Binder, *A Guide to Monte Carlo Simulations in Statistical Physics* (Cambridge University Press, Cambridge, 2000).

[162] B. Derrida and V. Hakim, *J. Phys. A* **29**, L589 (1996).

[163] P. L. Krapivsky, *J. Phys. A* **37**, 6917 (2004); arXiv:cond-mat/0405469.

Chapter 9: Coarsening

[164] J. S. Langer, An introduction to the kinetics of first-order phase transitions, in: *Solids Far From Equilibrium*, ed. C. Godrèche (Cambridge University Press, Cambridge, 1992).

[165] P. C. Hohenberg and B. I. Halperin, *Rev. Mod. Phys.* **49**, 435 (1977).

[166] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics* (Pergamon Press, Oxford, 1987).

[167] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953).

[168] M. E. Gage and R. S. Hamilton, *J. Differential Geom.* **23**, 69 (1986).

[169] M. A. Grayson, *J. Differential Geom.* **26**, 285 (1987).

[170] W. W. Mullins, *J. Appl. Phys.* **27**, 900 (1956).

[171] H. A. Schwarz, *Gesammelte Mathematische Abhandlungen*, Vol. 1 (Julius Springer, Berlin, 1890).

[172] A. Schoen, *Infinite Periodic Minimal Surfaces without Self-Intersection*, NASA Technical Note TN D-5541 (1970).

[173] H. Karcher, *Manuscripta Math.* **64**, 291 (1989).

[174] S. Fujimori and M. Weber, *Manuscripta Math.* **129**, 29 (2009).

[175] S. Coleman, *Aspects of Symmetry* (Cambridge University Press, Cambridge, 1988).

[176] R. Rajaraman, *Solitons and Instantons* (North-Holland, Amsterdam, 1989).

[177] N. Manton and P. Sutcliffe, *Topological Solitons* (Cambridge University Press, Cambridge, 2004).

[178] G. H. Derrick, *J. Math. Phys.* **5**, 1252 (1964).

[179] M. Zapotocky and W. Zakrzewski, *Phys. Rev. E* **51**, R5189 (1995).

[180] I. M. Lifshitz and V. V. Slyozov, *Zh. Eksp. Teor. Fiz.* **35**, 479 (1959).

[181] C. Wagner, *Z. Elektrochem.* **65**, 581 (1961).

[182] T. Nagai and K. Kawasaki, *Physica A* **120**, 587 (1983); K. Kawasaki and T. Nagai, *Physica A* **121**, 175 (1983).

[183] A. J. Bray and B. Derrida, *Phys. Rev. E* **51**, 1633 (1995); arXiv:cond-mat/9411037.

[184] A. J. Bray, B. Derrida, and C. Godrèche, *Europhys. Lett.* **27** 175 (1994).

[185] K. Sekimoto, *Int. J. Mod. Phys. B* **5**, 1843 (1991).

Chapter 10: Disorder

[186] B. Derrida and E. Gardner, *J. Physique* **47**, 959 (1986).

[187] P. L. Krapivsky, *J. Physique I* **1**, 1013 (1991).

[188] S. Redner, *Am. J. Phys.* **58**, 267 (1990).

[189] S. N. Majumdar and O. C. Martin, *Phys. Rev. E* **74**, 061112 (2006); arXiv:cond-mat/0609735.

[190] F. Hivert, S. Nechaev, G. Oshanin, and O. Vasilyev, *J. Stat. Phys.* **126**, 243 (2007); arXiv:cond-mat/0509584.

[191] S. Carmi, P. L. Krapivsky, and D. ben-Avraham, *Phys. Rev. E* **78**, 066111 (2008); arXiv:0808.0388.

[192] C. M. Newman and D. L. Stein, *Phys. Rev. E* **60** 5244 (1999); arXiv:cond-mat/9908455.

[193] J.-P. Bouchaud, A. Comtet, A. Georges, and P. Le Doussal, *Ann. Phys.* **201**, 285 (1990).

[194] P. Le Doussal, C. Monthus, and D. S. Fisher, *Phys. Rev. E* **59**, 4795 (1999).

[195] D. ben-Avraham and S. Havlin, *Diffusion and Reactions in Fractals and Disordered Systems* (Cambridge University Press, Cambridge, 2000).

[196] S. Redner, *Physica D* **38**, 287 (1989).

[197] B. D. Hughes, *Random Walks and Random Environments*, Vol. 1, *Random Walks*, Vol. 2, *Random Environments* (Oxford University Press, Oxford, 1995, 1995).

[198] M. Kardar, *Statistical Physics of Fields* (Cambridge University Press, Cambridge, 2007).

Chapter 11: Hysteresis

[199] T. Tomé and M. J. de Oliveira, *Phys. Rev. A* **41**, 4251 (1990).

[200] M. Rao, H. R. Krishnamurthy, and R. Pandit, *Phys. Rev. B* **42**, 856 (1990).

[201] G. H. Goldsztein, F. Broner, and S. H. Strogatz, *SIAM J. Appl. Math.* **57**, 1163 (1997).

[202] N. Berglund and H. Kunz, *J. Phys. A* **32**, 15 (1999).

[203] S. H. Strogatz, *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering* (Addison-Wesley, Reading, MA, 1994).

[204] Y. Imry and S.-K. Ma, *Phys. Rev. Lett.* **39**, 1399 (1975).

[205] R. P. Feynman, R. B. Leighton, and M. Sands, *The Feynman Lectures on Physics*, Vol. II (Addison-Wesley, Reading, MA, 1964).

[206] *Science of Hysteresis*, eds G. Bertotti and I. Mayergoyz (Academic Press, New York, 2006). This book contains many useful reviews, particularly J. P. Sethna, K. A. Dahmen, and O. Perkovic, Random-field Ising models of hysteresis (arXiv:cond-mat/0406320); G. Durin and S. Zapperi, The Barkhausen effect (arXiv:cond-mat/0404512).

[207] J. P. Sethna, *Statistical Mechanics: Entropy, Order Parameters, and Complexity* (Oxford University Press, New York, 2006).

[208] D. Dhar, P. Shukla, and J. P. Sethna, *J. Phys. A* **30**, 5259 (1997).

[209] P. Shukla, *Phys. Rev. E* **62**, 4725 (2000); *Phys. Rev. E* **63**, 027102 (2001).

[210] F. Colaiori, A. Gabrielli, and S. Zapperi, *Phys. Rev. B* **65**, 224404 (2002).

[211] B. K. Chakrabarti and M. Acharyya, *Rev. Mod. Phys.* **71**, 847 (1999).

Chapter 12: Population dynamics

[212] J. D. Murray, *Mathematical Biology* (Springer, Berlin, 1989).

[213] G. F. Gause, *The Struggle for Existence* (Dover, New York, 2003).

[214] R. M. May and W. J. Leonard, *SIAM J. Appl. Math.* **29**, 243 (1975).

[215] B. Hayes, *American Scientist* **93**, 207 (2005).

[216] D. P. Maki and M. Thompson, *Mathematical Models and Applications* (Prentice-Hall, Englewood Cliffs, NJ, 1973).

[217] R. A. Fisher, *The Genetical Theory of Natural Selection* (Clarendon Press, Oxford, 1930).

[218] S. Wright, *Genetics* **16**, 97 (1931).

[219] J. Hein, M. H. Schierup, and C. Wiuf, *Gene Genealogies, Variation and Evolution: a Primer in Coalescent Theory* (Oxford University Press, Oxford, 2005).

[220] J. Wakeley, *Coalescent Theory: An Introduction* (Roberts & Co., Greenwood Village, CO, 2008).

[221] M. Serva, *J. Stat. Mech.* P07011 (2005); arXiv:q-bio/0503036.

[222] D. Simon and B. Derrida, *J. Stat. Mech.* P05002 (2006); arXiv:cond-mat/0601167.

[223] M. Delbrück, *J. Chem. Phys.* **8**, 120 (1940).

[224] D. G Kendall, *J. Roy. Statist. Soc. Ser. B* **11**, 230 (1949).

[225] D. A. McQuarrie, *J. Chem. Phys.* **38**, 433 (1963).

[226] D. A. McQuarrie, C. J. Jachimowski, and M. E. Russell, *J. Chem. Phys.* **40**, 2914 (1964).

[227] M. Assaf and B. Meerson, *Phys. Rev. E* **74**, 041115 (2006).

[228] M. I. Freidlin and A. D. Wentzell, *Random Perturbations of Dynamical Systems* (Springer, New York, 1984).

[229] N. T. J. Bailey, *The Mathematical Theory of Infectious Diseases* (Oxford University Press, Oxford, 1987).

[230] V. Elgart and A Kamenev, *Phys. Rev. E* **70**, 041106 (2004).

[231] C. R. Doering, K. V. Sargsyan, and L. M. Sander, *Multiscale Model. Simul.* **3**, 283 (2005).

[232] M. Assaf and B. Meerson, *Phys. Rev. E* **75**, 031122 (2007).

[233] M. Khasin and M. I. Dykman, arXiv:0904.1737.

Chapter 13: Diffusive reactions

[234] B. Duplantier, *Proc. Symp. Pure Math.* **72**, Part 2, pp. 365–482 (AMS, Providence, RI, 2004); arXiv:math-ph/0303034.

[235] B. P. Lee, *J. Phys. A* **27**, 2633 (1994).

[236] P. L. Krapivsky, *Phys. Rev. E* **49**, 3233 (1994); arXiv:cond-mat/9403041.

[237] B. Ya. Balagurov and V. G. Vaks, *Sov. Phys. JETP* **38**, 968 (1974).

[238] P. Grassberger and I. Procaccia, *J. Chem. Phys.* **77**, 6281 (1982).

[239] M. Donsker and S. R. S. Varadhan, *Comm. Pure Appl. Math.* **32**, 721 (1979).

[240] G. Polya and G. Szego, *Isoperimetric Inequalities in Mathematical Physics* (Princeton University Press, Princeton, NJ, 1951).

[241] I. Chavel, *Eigenvalues in Riemannian Geometry* (Academic Press, Orlando, FL, 1984).

[242] Ya. B. Zeldovich, *Zh. Tekh. Fiz.* **19**, 1199 (1949); Ya. B. Zeldovich and A. S. Mikhailov, *Sov. Phys. Usp.* **30**, 23 (1988).

[243] Ya. B. Zeldovich and A. A. Ovchinnikov, *Chem. Phys.* **28**, 215 (1978).

[244] D. Toussaint and F. Wilczek, *J. Chem. Phys.* **78**, 2642 (1983).

[245] A. A. Ovchinnikov, S. F. Timashev, and A. A. Belyi, *Kinetics of Diffusion Controlled Chemical Processes* (Nova Science, Hauppage, NY, 1989).

[246] S. Redner and F. Leyvraz, in: *Fractals and Disordered Systems*, Vol. II, eds A. Bunde and S. Havlin (Springer, Berlin, 1993).

[247] S. Redner, Scaling theories of diffusion-controlled and ballistically controlled bimolecular reactions, chap. 1 in [73].

[248] M. Bramson and J. L. Lebowitz, *J. Stat. Phys.* **65**, 941 (1991).

[249] D. ben-Avraham, *J. Chem. Phys.* **88**, 941 (1988).

[250] T. Oppelstrup, V. V. Bulatov, A. Donev, M. H. Kalos, G. H. Gilmer, and B. Sadigh, *Phys. Rev. E* **80**, 066701 (2009).

[251] Z. Rácz, *Phys. Rev. Lett.* **55**, 1707 (1985).

[252] J. L. Spouge, *Phys. Rev. Lett.* **60**, 871 (1988).

[253] B. R. Thomson, *J. Phys. A* **22**, 879 (1989).

[254] H. Takayasu, I. Nishikawa, and H. Tasaki, *Phys. Rev. A* **37**, 3110 (1988).

[255] A. Kolmogorov, I. Petrovsky, and N. Piscounov, *Moscow Univ. Bull. Math.* **1**, 1 (1937).

[256] R. A. Fisher, *Ann. Eugenics* **7**, 353 (1937).

[257] Z. Cheng, S. Redner, and F. Leyvraz, *Phys. Rev. Lett.* **62**, 2321 (1989).

[258] H. Hinrichsen, V. Rittenberg, and H. Simon, *J. Stat. Phys.* **86**, 1203 (1997); arXiv:cond-mat/9606088.

[259] L. Gálfi and Z. Rácz, *Phys. Rev. A* **38**, 3151 (1988).

[260] E. Ben-Naim and S. Redner, *J. Phys. A* **25**, L575 (1992).

Chapter 14: Complex networks

[261] H. Jeong, S. P. Mason, A.-L. Barabási, and Z. N. Oltvai, *Nature* **411**, 41 (2001).

[262] J. Moody, *Am. J. Sociol.* **107**, 679 (2001).

[263] P. Di Francesco, P. Ginsparg, and J. Zinn-Justin, *Phys. Rep.* **254**, 1 (1995); arXiv:hep-th/9306153.

[264] J. Ambjorn, B. Durhuus, and T. Jonsson, *Quantum Geometry: A Statistical Field Theory Approach* (Cambridge University Press, Cambridge, 2005).

[265] P. Erdős and A. Rényi, *Publ. Math. Inst. Hung. Acad. Sci.* **5**, 17 (1960).

[266] S. Janson, D. E. Knuth, T. Luczak, and B. Pittel, The birth of the giant component, *Random Struct. Algorithms* **4**, 233 (1993).

[267] B. Ballobás, *Random Graphs* (Academic Press, London, 1985).

[268] S. Janson, T. uczak, and A. Rucinski, *Random Graphs* (Wiley, New York, 2000).

[269] B. Bollobás, C. Borgs, J. T. Chayes, J. H. Kim, and D. B. Wilson, *Random Struct. Algorithms* **18**, 201 (2001).

[270] E. Ben-Naim and P. L. Krapivsky, *Phys. Rev. E* **71**, 026129 (2005).

[271] R. Otter, *Ann. Math.* **49**, 583 (1948).

[272] R. T. Smythe and H. Mahmoud, *Theory Prob. Math. Statist.* **51**, 1 (1995).

[273] G. U. Yule, *Phil. Trans. Roy. Soc. B* **213**, 21 (1925).

[274] H. A. Simon, *Biometrica* **42**, 425 (1955).

[275] R. Albert and A.-L. Barabási, *Rev. Mod. Phys.* **74**, 47 (2002).

[276] S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of Networks: From Biological Nets to the Internet and WWW* (Oxford University Press, Oxford, 2003).

[277] M. E. J. Newman, A.-L. Barabási, and D. J. Watts, *The Structure and Dynamics of Networks* (Princeton University Press, Princeton, NJ, 2006).

[278] A. Barrat, M. Barthélemy, and A. Vespignani, *Dynamical Processes on Complex Networks* (Cambridge University Press, Cambridge, 2008).

[279] P. L. Krapivsky and S. Redner, *Phys. Rev. E* **63**, 066123 (2001); arXiv:cond-mat/0011094.

[280] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, *Phys. Rev. E* **63**, 062101 (2001).

[281] P. L. Krapivsky and S. Redner, *J. Phys. A* **35**, 9517 (2002).

Index