

Boltzmann’s Dilemma: An Introduction to Statistical Mechanics via the Kac Ring*

Georg A. Gottwald[†]
Marcel Oliver[‡]

Abstract. The process of coarse-graining—here, in particular, of passing from a deterministic, simple, and time-reversible dynamics at the microscale to a typically irreversible description in terms of averaged quantities at the macroscale—is of fundamental importance in science and engineering. At the same time, it is often difficult to grasp and, if not interpreted correctly, implies seemingly paradoxical results. The kinetic theory of gases, historically the first and arguably most significant example, occupied physicists for the better part of the 19th century and continues to pose mathematical challenges to this day. In this paper, we describe the so-called Kac ring model, suggested by Mark Kac in 1956, which illustrates coarse-graining in a setting so simple that all aspects can be exposed both through elementary, explicit computation and through easy numerical simulation. In this setting, we explain a Boltzmannian “Stoßzahlansatz,” ensemble averages, the difference between ensemble averaged and “typical” system behavior, and the notion of entropy.

Key words. statistical mechanics, irreversibility, coarse-graining, entropy

AMS subject classifications. 82-01, 82C23, 82C40

DOI. 10.1137/070705799

I. Introduction. The concept that matter is composed of atoms and molecules is nowadays a truism, much like saying that the earth is a sphere and that it orbits around the sun. However, reconciling an atomistic description of nature with our everyday experience of how the world works is so much harder that any serious explanation is usually deferred to an advanced undergraduate or graduate course in statistical physics.

The point of difficulty is not that, at the atomic level, we should be using quantum mechanics, that relativistic effects may not be ignored, or that we should really start out with a description of the subatomic structure which itself is described by highly elaborate quantum theories. Rather, the issue is much more general and, from the point of view of constructing mathematical models for any kind of observable process, much more fundamental: all theories of microscopic physics are governed by laws that are invariant under reversal of time—the evolution of the system can be traced back into the past by the same evolution equation that governs the prediction into the future. Processes on macroscopic scales, on the other hand, are manifestly irreversible.

*Received by the editors October 19, 2007; accepted for publication (in revised form) March 6, 2009; published electronically August 5, 2009.

<http://www.siam.org/journals/sirev/51-3/70579.html>

[†]School of Mathematics and Statistics, University of Sydney, Sydney, NSW 2006, Australia (gottwald@maths.usyd.edu.au).

[‡]School of Engineering and Science, Jacobs University, 28759 Bremen, Germany (oliver@member.ams.org).

Fluids mix, but cannot be unmixed by the same process (with some notable exceptions [15, 8]). Steam engines use differences in temperature to do useful mechanical work; in doing so, heat flows from hot to cold, eventually equilibrates, and reduces the potential for doing further mechanical work. Particularly stark examples of irreversibility are birth, aging, and dying.

Scientists are also completely comfortable with mathematical modeling at the macroscale. Classical thermodynamics, reaction-diffusion equations, empirical laws for friction and drag, and many other equations of engineering and science provide extremely accurate quantitative descriptions of observable, irreversible phenomena without reference to the underlying microphysics. This begs an obvious question. Since we now have two presumably accurate, quantitative mathematical models of the same system, how does one embed into the other? And if such a relation can be established, how can it possibly be that one of these descriptions is reversible whereas the other is not?

The short answer, pioneered by Maxwell and Boltzmann in the context of a kinetic theory of gases, is that the macroscopic laws are a fundamentally statistical description—they represent the most probable behavior of the system. At the same time, most microscopic realizations of a macroscopic state remain close to the most probable behavior for a long, but finite interval of time. The long answer, which provides quantitative definiteness and mathematical rigor, is, in most cases, truly long and difficult. Nonetheless, there are a number of toy models, most notably the Ehrenfests' urn model [5] and Mark Kac's ring model [6], which provide a caricature of the underlying issues while being simple enough that they can be exposed through elementary and explicit computation. Many of the mathematical questions underlying the kinetic theory of gases remain open, though, and are central to making advances in an increasing number of applications of kinetic theory to the modeling of complex systems, e.g., in biology, climate, semiconductors, and economics [2].

These notes are structured as follows. After a brief historical introduction, section 3 describes the Kac ring model and the Kac ring analog of the Stoßzahlansatz in kinetic theory. In section 4, we define and compute ensemble averages. The conditions under which typical experiments will follow the averaged behavior are discussed, by computing the variance of the ensemble, in section 5. Section 6 presents the continuum limit of the Kac ring and section 7 the concept of entropy. The paper closes with a brief mention of the Ehrenfest model, an extended general summary and discussion, and a section in which we comment on some of our experiences when teaching lectures based on these ideas.

Our presentation follows, in part, the books by Dorfman [3] and Schulman [12]. We have not seen a detailed quantitative discussion of the variance and the continuum limit elsewhere, though they were certainly known to Kac [4]. The paper also contains a good number of exercises, from easy or review-type topics to more advanced open-ended questions which extend beyond a first acquaintance with the subject.

2. Historical Background. The questions we address here have historically been asked in light of the apparent conflict between classical thermodynamics and mechanics which followed the chemists' realization that gases are composed of discrete, apparently indivisible molecules. We shall give a brief account of this background, not to strive for completeness and accuracy (that would require a more major endeavor than anything we attempt here), but to convey the sense that the concepts ultimately provided by Maxwell and Boltzmann constituted a major breakthrough in scientific thought. For more historic detail, see [10, 13, 17] and the references therein.

Up until the end of the 18th century, the mathematical theories of thermodynamics and mechanics were studied largely in parallel. When Isaac Newton (1643–1727) wrote down the foundations of classical mechanics and Robert Boyle (1627–1691) announced his law which relates the volume of a gas to its pressure, there was little concrete evidence that the two were related at all. Still, the idea that heat is a consequence of the motion of atomic particles dates back to at least this era and is discussed in the works of Boyle, Newton, and later of Daniel Bernoulli (1700–1782), who, in 1738, first formulated a quantitative kinetic theory. This work, however, was largely forgotten for another century.

The issue acquired new urgency when John Dalton (1766–1844) and others laid the foundations for the modern atomic theory of matter and Amedeo Avogadro (1776–1856) postulated that the number of molecules per volume is a function only of temperature and pressure, independent of the chemical composition of a gas. A number of scientists, most notably Herapath, Waterson, Joule, and Clausius (who is also credited with stating the second law of thermodynamics in terms of entropy), worked on kinetic descriptions of gases.

The real breakthrough in the development of kinetic theory, however, is due to James Clerk Maxwell (1831–1879) and Ludwig Boltzmann (1844–1906). Maxwell was the first to look at gases from a probabilistic point of view and he computed the equilibrium probability distribution of molecular velocities (which, in fact, is a manifestation of the central limit theorem). In a seminal 1867 paper, Maxwell introduced an evolution equation for (moments of) the probability density of finding a molecule of given velocity at a given location in space, which he used to derive equations for the macroscopic dynamics of gases. The crucial underlying assumption, which Boltzmann later called the *Stoßzahlansatz* (which literally translates as “collision number hypothesis” and is often referred to as the *molecular chaos assumption*), is the statistical independence of the probability densities of colliding molecules (i.e., the model has no memory of trajectories of individual molecules).

Among Boltzmann’s most prominent contributions is an 1872 paper in which he wrote out a reformulation of Maxwell’s kinetic equation, which today is referred to as the *Boltzmann equation*, from which he deduced that there is a quantity H which is nonincreasing along solution trajectories of the Boltzmann equation and, in particular, remains constant for Maxwell’s equilibrium distribution. Boltzmann equated the negative of his H -function with Clausius’ thermodynamic entropy and thus claimed to have proved the second law of thermodynamics via statistical mechanics.

Boltzmann’s work faced severe criticism, essentially on two levels. His contemporaries asked, almost immediately after publication, how a microscopically time-reversible and recurrent system can give rise to a macroscopic description which is neither. The two famous objections became known as *Loschmidt’s paradox* and *Zermelo’s paradox*. Josef Loschmidt (1821–1895) and others pointed out that system trajectories in classical mechanics, the starting point of Boltzmann’s derivation, are deterministic and time-reversible—see Exercise 1. Therefore, the reasoning goes, Boltzmann’s result, namely, the manifestly irreversible H -theorem, cannot possibly be correct. (In fact, Maxwell himself had already addressed and essentially resolved the reversibility paradox several years earlier.) Ernst Zermelo (1871–1953) gave a new, general proof of Poincaré’s recurrence theorem which showed that the theorem is indeed applicable to the situation considered by Boltzmann [18]. Poincaré’s theorem states that a dynamical system with constant energy in a compact phase space must eventually return to its initial state within arbitrary precision for almost all initial data [9]. Thus, surely, H cannot always decrease but must also increase as the

system returns to its initial configuration. While Zermelo's proof is based on abstract measure theory, the issue can be easily understood in a finite setting. It hinges on the property of reversibility—see Exercise 2.

Exercise 1. Newton's second law of mechanics for a particle of mass m situated at position $x(t)$ moving with velocity $v(t)$ and subject to a force $F(x(t))$ can be written

$$(1a) \quad \frac{dx}{dt} = v,$$

$$(1b) \quad m \frac{dv}{dt} = F(x(t)).$$

Show that the particle satisfies the same equation with t replaced by the reversed time $s = -t$ and v replaced by $u = -v$.

Exercise 2. Show that in a time-discrete, reversible system with a finite number of states, any orbit must return to its initial state after a finite number of steps.

Remark 1. Reversibility applies not only to classical mechanics, but, in a more general sense, to all physical laws governing the fundamental forces of nature. The resulting apparent discrepancy with our macroscopic experience of the world is often considered to be a major philosophical concern. We take the view that the issue of coarse-graining as a modeling paradigm is largely a mathematical question. Once the operational issues are understood, the urgency of philosophical debate is greatly diminished. Similar sentiments have, for example, been expressed in [11].

The second strand of objections to Boltzmann's approach concerns the justification of the Stoßzahlansatz and of the subsequent reduction of the macroscopic dynamics to an evolution equation for a one-particle probability distribution function. While it is now generally acknowledged that Boltzmann did not answer this question adequately and thus did not provide a proof of the second law of thermodynamics—in fact, the Boltzmann entropy equals the thermodynamic entropy only in the ideal gas limit—the lasting impact of Boltzmann's contribution lies in the extension of the notion of entropy to nonequilibrium systems [17]. More generally, this observation highlights the differences between what has been called the *master equation* and the *Liouville equation* approaches to statistical physics (see, for example, [7]). In the former, a Boltzmannian Stoßzahlansatz simplifies the model and evolves a macrostate (a technically precise definition would be based on the Markov property of the resulting stochastic process). In the latter, the dynamics is given by the full evolution at the microscopic level; macroscopic information is extracted a posteriori by computing relevant averages. Kac's ring model, which we explain in this paper, is an ingeniously simple system in which both the master equation approach (section 3) and the Liouville equation approach (section 4) can be carried out explicitly. As we shall see, the Stoßzahlansatz loses information but provides a valid description under certain conditions. Generally, however, coarse-graining and time evolution do not commute. While the Liouville equation approach provides, in principle, a correct description, it may be theoretically and computationally intractable; typically, it is successful only at equilibrium. We will return to this point in our final discussion; see section 9.

3. The Kac Ring Model. The *Kac ring* is a simple, explicitly solvable model which illustrates the process of passing from a microscopic, time-reversible description to a macroscopic, thermodynamic description. In this model, N sites are arranged around a circle, forming a one-dimensional *periodic lattice*; neighboring sites are joined by an *edge*. Each site is occupied by either a black ball or a white ball. Moreover, $n < N$ of the edges carry a marker; see Figure 1.

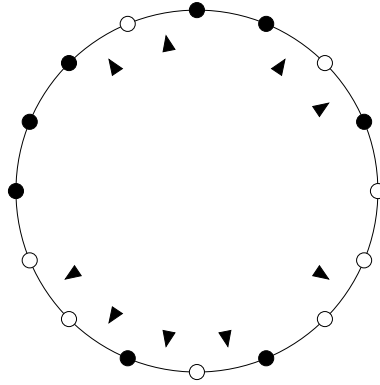


Fig. 1 A Kac ring with $N = 16$ lattice sites and $n = 9$ markers.

The system evolves on a discrete set of clock ticks $t \in \mathbb{Z}$ from state t to state $t + 1$ as follows. Each ball moves to the clockwise neighboring site. When a ball passes a marker, its color changes.

The dynamics of the Kac ring is time-reversible and has recurrence. When the direction of movement along the ring is reversed, the balls retrace their past color sequence without change to the “laws of physics.” Moreover, after N clock ticks, each ball has reached its initial site and changed color n times. Thus, if n is even, the initial state recurs; if n is odd, it takes at most $2N$ clock ticks for the initial state to recur.

Let $B(t)$ denote the total number of black balls and $b(t)$ the number of black balls just in front of a marker; let $W(t)$ denote the number of white balls and $w(t)$ the number of white balls in front of a marker. Then

$$(2) \quad B(t + 1) = B(t) + w(t) - b(t)$$

and, similarly,

$$(3) \quad W(t + 1) = W(t) + b(t) - w(t).$$

We will study the behavior of $\Delta(t) = B(t) - W(t)$. Clearly,

$$(4) \quad \Delta(t + 1) = B(t + 1) - W(t + 1) = \Delta(t) + 2w(t) - 2b(t).$$

Note that W , B , and Δ are *macroscopic* quantities, describing a global feature of the system state; w and b , on the other hand, contain local information about individual sites—they cannot be computed without knowing the location of each marker and the color of the ball at every site. A key feature of this system is that the evolution of the global quantities is not computable from only macroscopic state information. In other words, it is not possible to eliminate b and w from (2)–(4). This is known as the *closure problem*.

When the markers are distributed at random, the probability that a particular site is occupied by a marker is given by

$$(5) \quad \mu \equiv \frac{n}{N} = \frac{b}{B} = \frac{w}{W}.$$

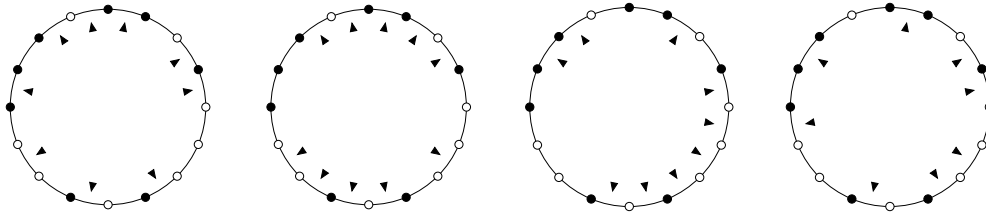


Fig. 2 An ensemble of Kac rings in their initial state. The configuration of black and white balls is fixed across the ensemble. Each edge carries a marker with probability μ .

For an actual realization of the Kac ring these relations will generally not be satisfied. However, by assuming that they hold anyway, we can overcome the closure problem. This assumption is the analogue of Maxwell and Boltzmann's Stoßzahlansatz. It effectively disregards the history of the system evolution: there is no memory of where the balls originated and which markers they passed up to time t . However, we hope that this assumption represents, in some sense, the typical behavior of large-sized rings.

Under this Stoßzahlansatz, (4) becomes

$$(6) \quad \Delta(t+1) = \Delta(t) + 2\mu W(t) - 2\mu B(t) = (1 - 2\mu) \Delta(t).$$

This recurrence immediately yields

$$(7) \quad \Delta(t) = (1 - 2\mu)^t \Delta(0).$$

For our model, (7) takes the role of the *Boltzmann equation* in the kinetic theory of gases.

Clearly, this equation cannot describe the dynamics of one particular ring exactly. For instance, $\Delta(t)$ is generically not an integer anymore. Moreover, since $\mu < 1$, we see that $\Delta(t) \rightarrow 0$ as $t \rightarrow \infty$. Contrary to what we know about the microscopic dynamics, the magnitude of Δ in (7) is monotonically decreasing and therefore not time-reversible—we have an instance of *Loschmidt's paradox*. Moreover, the initial state cannot recur, again in contrast to the microscopic dynamics which has a recurrence time of at most $2N$ —we also have an instance of *Zermelo's paradox*.

Exercise 3. Explain the identities in (5).

Exercise 4. A Kac ring with N sites is initially occupied by B black and W white balls. The markers are distributed at random, each edge carrying a marker with probability μ . Now consider a single turn of the ring. Give an expression for the probability that the ring is occupied by only white balls at $t = 1$. How does this probability behave for large N ?

4. Ensemble Averages. Our task is to give a meaning to the macroscopic evolution equation (7) on the basis of the microscopic dynamics. Boltzmann suggested that the macroscopic law can only be valid in a statistical sense, referring to the most probable behavior of a member in a large ensemble of systems rather than to the exact behavior of any member of the ensemble. For the Kac ring, this notion is easily made precise and results in explicitly computable predictions.

By an ensemble of Kac rings we mean a collection of rings with the same number of sites as is depicted in Figure 2. Each member of the ensemble has the same initial configuration of black and white balls. The markers, however, are placed at random

such that the probability that any one edge is occupied by a marker equals μ . Let X denote some function of the configuration of markers and X_j denote the value of X for the j th member of the ensemble. Then the *ensemble average* $\langle X \rangle$ is defined as the arithmetic mean of X over a large number of realizations, i.e.,

$$(8) \quad \langle X \rangle = \lim_{M \rightarrow \infty} \frac{1}{M} \sum_{j=1}^M X_j.$$

In the language of (finite) probability, each particular configuration of markers is referred to as an *outcome* from among the *sample space* S of all possible configurations of markers. The process of choosing a random configuration of markers is called a *trial*. It is always assumed that trials are *independent*, i.e., that the result of one trial does not depend on any past trials. We now recognize X as a function $X: S \rightarrow \mathbb{R}$; any such function is referred to as a *random variable*. Thus, the ensemble average is nothing but the *expected value* of a random variable, and it can be computed as follows. As the system is finite, X will take one of x_1, \dots, x_I possible values (“macrostates”) with corresponding probabilities p_1, \dots, p_I . Then

$$(9) \quad \langle X \rangle = \sum_{i=1}^I p_i x_i.$$

The identification of (8) and (9) is due to the definition of the probability p_i of the event $\{X = x_i\}$ as its *relative frequency of occurrence in a large number of trials*. (In general, the term *event* refers to a subset of the sample space.)

Exercise 5. Verify the identity of (8) and (9) explicitly for the following simple case. Consider an ensemble of systems which can be in one of two states. State 1 occurs with probability p_1 , where X takes the value $x_1 = 1$; state 2 occurs with probability $p_2 = 1 - p_1$, where X takes the value $x_2 = 0$.

Returning to the Kac ring, we formalize the microscopic laws of evolution as follows. Let $\chi_i(t)$ denote the color of the ball occupying the i th lattice site at time t , with value 1 representing black and value -1 representing white. Further, let $m_i = -1$ denote the presence and $m_i = 1$ denote the absence of a marker on the edge connecting sites i and $i + 1$. Then the recurrence relation for stepping from t to $t + 1$ reads

$$(10) \quad \chi_{i+1}(t+1) = m_i \chi_i(t).$$

This expression makes sense for any $i, t \in \mathbb{Z}$ if we identify $\chi_0 \equiv \chi_N$, $\chi_1 \equiv \chi_{N+1}$, and so on—the lattice is periodic with period N . Similarly, we have $m_0 \equiv m_N$, $m_1 \equiv m_{N+1}$, and so on. Then

$$(11) \quad \Delta(t) = \sum_{i=1}^N \chi_i(t) = \sum_{i=1}^N m_{i-1} \chi_{i-1}(t-1) = \dots = \sum_{i=1}^N m_{i-1} m_{i-2} \dots m_{i-t} \chi_{i-t}(0).$$

Exercise 6. Verify, using (11), that $\Delta(2N) = \Delta(0)$.

Equipped with the above definition and recipe for computing the ensemble average, we wish to compute the evolution of $\langle \Delta(t) \rangle$. Since averaging involves taking sums and therefore distributes over sums, and since only the marker positions, not the initial configuration of balls, may differ across the ensemble, we can pull the sum

and the $\chi_i(0)$ out of the average and obtain

$$(12) \quad \langle \Delta(t) \rangle = \sum_{i=1}^N \langle m_{i-1} m_{i-2} \cdots m_{i-t} \rangle \chi_{i-t}(0).$$

Since all lattice edges have equal probability of carrying a marker, the average in (12) must be invariant under index shifts. In particular, $\langle m_{i-1} m_{i-2} \cdots m_{i-t} \rangle = \langle m_1 m_2 \cdots m_t \rangle$, so that

$$(13) \quad \langle \Delta(t) \rangle = \langle m_1 m_2 \cdots m_t \rangle \sum_{i=1}^N \chi_{i-t}(0) = \langle m_1 m_2 \cdots m_t \rangle \Delta(0).$$

Our remaining task is to find an explicit expression for $\langle m_1 m_2 \cdots m_t \rangle$, a quantity which depends only on the distribution of the markers, not on the balls. We distinguish two cases.

When $0 \leq t < N$, there are no periodicities—all factors m_1, \dots, m_t are independent. The value of the product is 1 for an even number of markers and -1 for an odd number of markers. Thus, (9) takes the form

$$(14) \quad \langle m_1 m_2 \cdots m_t \rangle = \sum_{j=0}^t (-1)^j p_j(t),$$

where $p_j(t)$ denotes the probability of finding j markers on t consecutive edges. The markers follow a *binomial distribution*, so that

$$(15) \quad p_j(t) = \mu^j (1 - \mu)^{t-j} \binom{t}{j},$$

and, using the binomial theorem,

$$(16) \quad \langle m_1 m_2 \cdots m_t \rangle = \sum_{j=0}^t \binom{t}{j} (-\mu)^j (1 - \mu)^{t-j} = (1 - 2\mu)^t.$$

Inserting this expression into (13), we obtain

$$(17) \quad \langle \Delta(t) \rangle = (1 - 2\mu)^t \Delta(0),$$

the same expression (7) we obtained through our initial “molecular chaos assumption.”

This result is encouraging, because it shows that the relatively crude Stoßzahlansatz of section 3 may be related to the average over a statistical ensemble. In general, however, one cannot expect exact identity. Indeed, even for the Kac ring, the following computation shows that when $t > N$, the two concepts diverge.

Exercise 7. Derive (15). Proceed in two steps. First, determine the probability of finding a marker on each of the first j edges and no marker on the next $t - j$ edges. Second, correct for the resulting undercount by considering all possible distinct permutations of these edges.

When $N \leq t < 2N$, balls may pass some markers twice, and we have to explicitly account for these periodicities:

$$(18) \quad \langle m_1 \cdots m_t \rangle = \langle m_{t+1} \cdots m_{2N} \rangle = \langle m_1 \cdots m_{2N-t} \rangle.$$

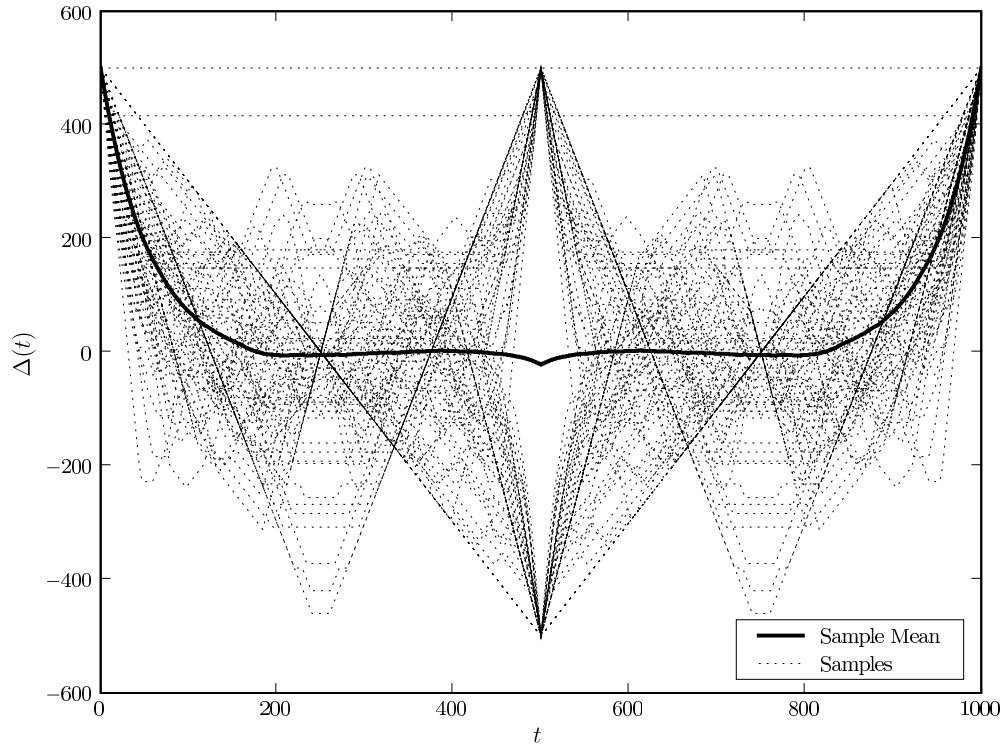


Fig. 3 The evolution of $\Delta(t)$ for an ensemble of $M = 400$ Kac rings with $N = 500$ sites which are initially occupied by black balls over the full recurrence time $t = 2N$. Edges carry markers with probability $\mu = 0.009$.

The first equality is a consequence of the N -periodicity of the lattice, namely, that $m_i = m_{N+i}$, which implies that $m_1 m_2 \cdots m_{2N} = 1$. The second equality is due again to the invariance of the average under an index shift. We may thus follow the argument leading from (14) to (17) with $2N - t$ in place of t , to finally obtain

$$(19) \quad \langle \Delta(t) \rangle = (1 - 2\mu)^{2N-t} \Delta(0).$$

As the exponent on the right-hand side is negative on the interval $N \leq t \leq 2N$, the ensemble average $\langle \Delta(t) \rangle$ increases on this interval and, in particular, recurs to its initial value for $t = 2N$. This behavior is called *anti-Boltzmann*.

Figure 3 shows a simulation of an ensemble of Kac rings. Clearly visible is the recurrence at $t = 2N$. The sample mean compares well with the predicted ensemble average; see (17) and (19). The theoretical predictions are not explicitly drawn, but they would not visibly depart from the sample mean except near the half-recurrence time $t = N$, when each individual trajectory either recurs or arrives at the negative of its initial value. The half-recurrence time is also the time of maximal variance, as we will discuss in section 5, so that the size of our sample is simply too small to get a good experimental prediction of the mean. Figure 3 also shows that some realizations have large departures from the mean while most stay close, at least for times $t \ll N$. We shall make this observation more precise in section 6. Finally, note that the expected number of markers per ring in Figure 3 is 4.5. Thus, with high

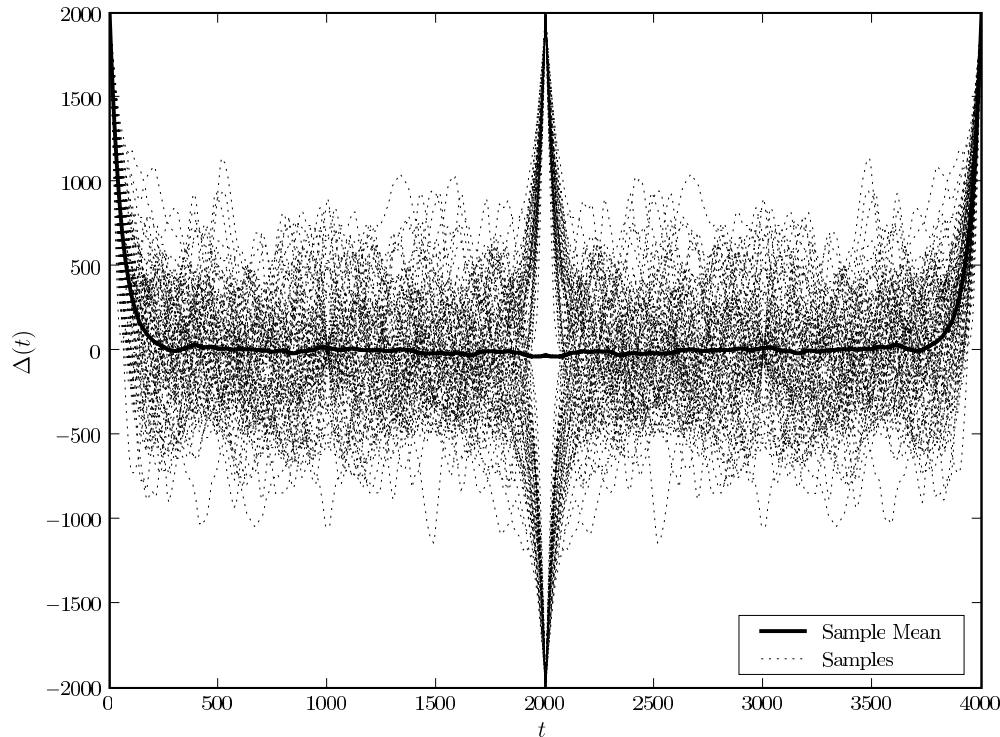


Fig. 4 Ensemble of larger sized Kac rings with $N = 2000$; all other parameters are as in Figure 3.

probability, the ensemble contains rings with zero, one, and two markers. Trajectories corresponding to each of these cases are clearly visible. (Can you identify them?)

Figure 4 shows the behavior of an ensemble of larger-sized Kac rings with all other parameters kept fixed. For larger rings, the probability of having a ring without a marker or with just a few markers becomes negligible. Thus, the sample shown does not contain trajectories corresponding to almost marker-free members of the ensemble.

5. Is Average Behavior Typical? In the previous section, we have shown that the Stoßzahlansatz leads to a macroscopic equation that represents the averaged behavior of an ensemble of Kac rings for times $t \leq N$. This, however, does not imply that the ensemble average represents in some way “typical” members of the ensemble, or that it is even close to any individual system trajectory. For example, at the half-recurrence time $t = N$, each ball is back at its initial position with a possible global change of color whenever the total number of markers is odd, so that $\Delta(N) = \pm\Delta(0)$ while, by (17), $\langle\Delta(N)\rangle$ is close to zero. For small t , on the other hand, most members of the ensemble stay close to the sample mean. Both of these regimes are clearly visible in Figures 3 and 4. How can we quantify this observed behavior?

We will answer this question by estimating the *variance* of the ensemble as a function of t , which is a measure of how much the individual members of the ensemble disperse about the mean. More precisely,

$$(20) \quad \text{Var}[\Delta(t)] = \langle(\Delta(t) - \langle\Delta(t)\rangle)^2\rangle.$$

In particular, the variance is zero if all members coincide with the mean and $\text{Var}[\Delta] = d^2$ if all members are situated at distance $\pm d$ away from the mean. An easy computation gives an alternative expression, more amenable to computation,

$$(21) \quad \text{Var}[\Delta(t)] = \langle \Delta^2(t) \rangle - \langle \Delta(t) \rangle^2.$$

The estimation of the variance follows essentially the same pattern as the computation of the ensemble mean in section 4. By analogy with (11),

$$(22) \quad \begin{aligned} \Delta^2(t) &= \sum_{i,j=1}^N \chi_i(t) \chi_j(t) \\ &= \sum_{i,j=1}^N m_{i-1} \cdots m_{i-t} m_{j-1} \cdots m_{j-t} \chi_{i-t}(0) \chi_{j-t}(0) \\ &= \sum_{k=-N/2}^{N/2-1} \sum_{i=1}^N m_{i-1} \cdots m_{i-t} m_{i-1+k} \cdots m_{i-t+k} \chi_{i-t}(0) \chi_{i-t+k}(0), \end{aligned}$$

where we have reindexed one of the sums writing $j = i+k$, noting again the periodicity of the lattice. Moreover, for simplicity, we assume that N is even so that $N/2$ is an integer. Then taking averages and noting, once more, that the average over products of the m_i is invariant under index shifts, we find that

$$(23) \quad \langle \Delta^2(t) \rangle = \sum_{k=-N/2}^{N/2-1} \langle m_t \cdots m_1 m_{t+k} \cdots m_{1+k} \rangle \sum_{i=1}^N \chi_{i-t}(0) \chi_{i-t+k}(0).$$

To proceed, we have to answer the following question: How many *independent* terms are there within the averaging brackets? If $t < N/2$, which we shall assume throughout, there can certainly be no more than $2t$ independent factors. For small values of $|k| < t$, however, there are only $|k|$ terms in the first group of factors which do not recur in the second group of factors, and vice versa, for a total of $2|k|$ independent factors. We conclude, repeating the argument that previously led us from (14) to (17), that

$$(24) \quad \langle m_t \cdots m_1 m_{t+k} \cdots m_{1+k} \rangle = (1 - 2\mu)^{2\min\{|k|,t\}}.$$

Hence,

$$(25) \quad \begin{aligned} \langle \Delta^2(t) \rangle &= \sum_{k=-N/2}^{N/2-1} (1 - 2\mu)^{2\min\{|k|,t\}} \sum_{i=1}^N \chi_{i-t}(0) \chi_{i-t+k}(0) \\ &= (1 - 2\mu)^{2t} \sum_{k=-N/2}^{N/2-1} \sum_{i=1}^N \chi_{i-t}(0) \chi_{i-t+k}(0) \\ &\quad + \sum_{k=1-t}^{t-1} \left((1 - 2\mu)^{2|k|} - (1 - 2\mu)^{2t} \right) \sum_{i=1}^N \chi_{i-t}(0) \chi_{i-t+k}(0) \\ &= (1 - 2\mu)^{2t} \Delta^2(0) \\ &\quad + \sum_{k=1-t}^{t-1} \left((1 - 2\mu)^{2|k|} - (1 - 2\mu)^{2t} \right) \sum_{i=1}^N \chi_{i-t}(0) \chi_{i-t+k}(0). \end{aligned}$$

The first term on the right, due to (17), equals $\Delta^2(t)$, so that, using the definition of the variance and the fact that the final sum of (25) has N terms of unit modulus,

$$\begin{aligned}
 \text{Var}[\Delta(t)] &\leq N \sum_{k=1-t}^{t-1} \left((1-2\mu)^{2|k|} - (1-2\mu)^{2t} \right) \\
 &= N \left(2 \frac{1 - (1-2\mu)^{2t}}{1 - (1-2\mu)^2} - 1 - (2t-1)(1-2\mu)^{2t} \right) \\
 (26) \quad &= N \left(\frac{1 - (1-2\mu)^{2t}}{2\mu(1-\mu)} - 1 - (2t-1)(1-2\mu)^{2t} \right).
 \end{aligned}$$

Exercise 8. Show that the bound (26) for $\text{Var}[\Delta(t)]$ is strictly increasing on $0 \leq t \leq N/2$ provided that $0 < \mu < \frac{1}{2}$.

Exercise 9. Show that $\text{Var}[\Delta(N)] = (1 - (1-2\mu)^{2N}) \Delta^2(0)$. Describe the behavior for large N .

Exercise 10. Repeat the analysis of this section on the interval $N/2 \leq t \leq N$. You should find that

$$(27) \quad \text{Var}[\Delta(t)] \leq N \left(\frac{1 - (1-2\mu)^{2(N-t)}}{2\mu(1-\mu)} - 1 + (2t-N+1)(1-2\mu)^{2(N-t)} - N(1-2\mu)^{2t} \right).$$

Remark 2. The bounds on the variance given by (26) and (27) hold with equality when $|\Delta| = N$ initially. For this reason, the simulations in Figures 3–5 were initialized in that manner.

The most important consequence we can draw from (26) and (27) is that the variance scales like N , and so the standard deviation scales like \sqrt{N} as long as we remain some distance away from the half-recurrence time $t = N$. This behavior indicates that as N gets large, the tube about $\langle \Delta(t) \rangle$ with a width of one standard deviation as depicted in Figure 5 becomes narrow relative to $\Delta_{\max} = N$. Thus, for short times and large N we can conclude that average behavior is typical. In the following section we shall make this asymptotic regime more precise.

6. Continuum Limit. So far, everything we have talked about was fully discrete: a finite, integer number of sites and markers and integer time t . The relative distances between neighboring discrete values for W , B , and Δ , however, decrease as the size of the Kac ring increases. In addition, we should by now have gained a sense that large rings are “more typical” than smaller rings.

In the following, we will show how to pass to the “large ring limit” in a rigorous and precisely defined sense. To begin, we note that simply letting $N \rightarrow \infty$ in the above computation will not work—everything will simply diverge in an uncontrolled manner. Therefore, the key is to identify quantities which neither diverge nor go to zero and can thus carry nontrivial information about the system behavior into the limit.

The first of such quantities is almost obvious. What carries meaning is not the absolute difference between the numbers of black and white balls Δ , but its magnitude *relative* to the total system size. We thus define

$$(28) \quad \delta = \frac{\Delta}{N},$$

a measure of the grayness of the ring when looked at from a great distance. The grayness ranges from $\delta = -1$, corresponding to complete whiteness, to $\delta = 1$, corresponding to complete blackness, independent of the system size N . Although for a

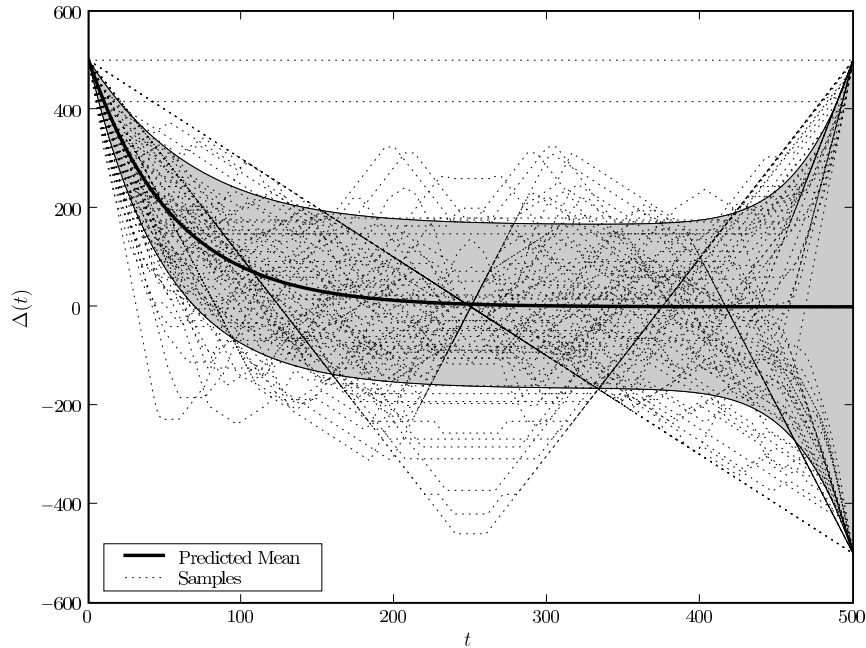


Fig. 5 Magnified view of the time window $0 \leq t \leq N$, where the solution depicted in Figure 3 has “Boltzmann behavior.” Also depicted is the neighborhood with the radius of one standard deviation (the square root of the variance) about the predicted ensemble mean as given by (26) and Exercise 10.

ring of a fixed size, δ takes only a discrete set of values, every real $\delta \in [-1, 1]$ can be approximated arbitrarily closely by a state of a finite Kac ring of sufficiently large size. In terms of δ , (17) and (26) read

$$(29) \quad \langle \delta(t) \rangle = (1 - 2\mu)^t \delta(0)$$

and

$$(30) \quad \text{Var}[\delta(t)] \leq \frac{1}{N} \left(\frac{1}{2\mu(1-\mu)} - 1 \right),$$

respectively.

Second, we want the system within one unit of macroscopic time to be affected by very many steps of the underlying microscopic Kac ring dynamics, in much the same way that the macroscopic grayness is affected by very many sites. This is achieved by introducing a macroscopic time variable τ , which relates to microscopic time t via a scaling law of the form

$$(31) \quad \tau = \frac{t}{N^\alpha}$$

for some exponent $\alpha > 0$.

Third, the behavior of the system within one unit of macroscopic time should be nontrivial as $N \rightarrow \infty$. Substituting (31) into (29), we see that it is necessary to have

$\mu \rightarrow 0$ in this limit. To be definite, we set

$$(32) \quad \mu = \frac{1}{2N^\beta}$$

for some exponent $\beta > 0$. The prefactor $1/2$ is for convenience, as we shall see below, and does not affect the nature of the result. We shall also require that $\beta < 1$, for otherwise there would be, on average, less than one marker per ring so that, in the limit, most realizations would be uninteresting. With $\beta \in (0, 1)$, the scaling law (32) expresses that the average number of markers goes to infinity, but at a rate less than the rate at which the size of the ring diverges.

Plugging the scaling assumptions into (29), we find that

$$(33) \quad (1 - 2\mu)^t = \left(1 - \frac{1}{N^\beta}\right)^{\tau N^\alpha} \rightarrow \begin{cases} 0 & \text{if } \beta < \alpha, \\ e^{-\tau} & \text{if } \beta = \alpha, \\ 1 & \text{if } \beta > \alpha, \end{cases}$$

as $N \rightarrow \infty$ for any fixed $\tau > 0$. Hence, the condition $\alpha = \beta$ is necessary for obtaining a nontrivial large system limit. Under this assumption, the limit Kac ring dynamics becomes

$$(34) \quad \langle \delta(\tau) \rangle = e^{-\tau} \delta(0).$$

Moreover,

$$(35) \quad \text{Var}[\delta] \leq \frac{1}{N} \frac{1}{2\mu(1-\mu)} \sim \frac{1}{N} \frac{1}{2\mu} = N^{\beta-1}$$

so that $\text{Var}[\delta] \rightarrow 0$ as $N \rightarrow \infty$ (recall that $\beta < 1$). (Two sequences a_k and b_k are said to be *asymptotic*, symbolically written $a_k \sim b_k$, if $\lim_{k \rightarrow \infty} a_k/b_k = 1$.) This proves that, in the limit, almost all Kac rings follow the ensemble averaged dynamics; the macroscopic equation (34) describes the macroscopically observable behavior of almost any realization. Figure 6 illustrates the relation between scaled and unscaled variables and the resulting limiting behavior of the ensemble.

We finally remark that the scaling laws for t and μ may look arbitrary. This is correct in the sense that such scalings are generally outside the scope of the fundamental laws of physics and lack uniqueness in any strict mathematical sense. It is rather up to the ingenuity of the modeler to come up with a scaling which induces a mathematically tractable and well-behaved limit—as in our example above—and, when modeling real-world systems, is consistent with the relevant physical parameters.

7. Entropy. The term *entropy* was coined by Robert Clausius (1822–1888) in the 1860s, at the time as a purely macroscopic concept without anticipation of an underlying probabilistic origin. It arises in the mathematical description of a thermodynamical process—a closed system undergoing adiabatic transformations, to be precise—as a quantity which can only increase in time.

In popular accounts of statistical thermodynamics, entropy is often described as a measure of “disorder.” This notion, however, is ill-defined and potentially misleading. Let us clarify the issue using a very simple example: a deck of N cards. Assume that the cards are all ordered in a well-defined way; for example, first all the hearts, then diamonds, then clubs, then spades, all of them in ascending order. Now shuffle the deck of cards. Is the resulting state of the deck more disordered than the initial one?

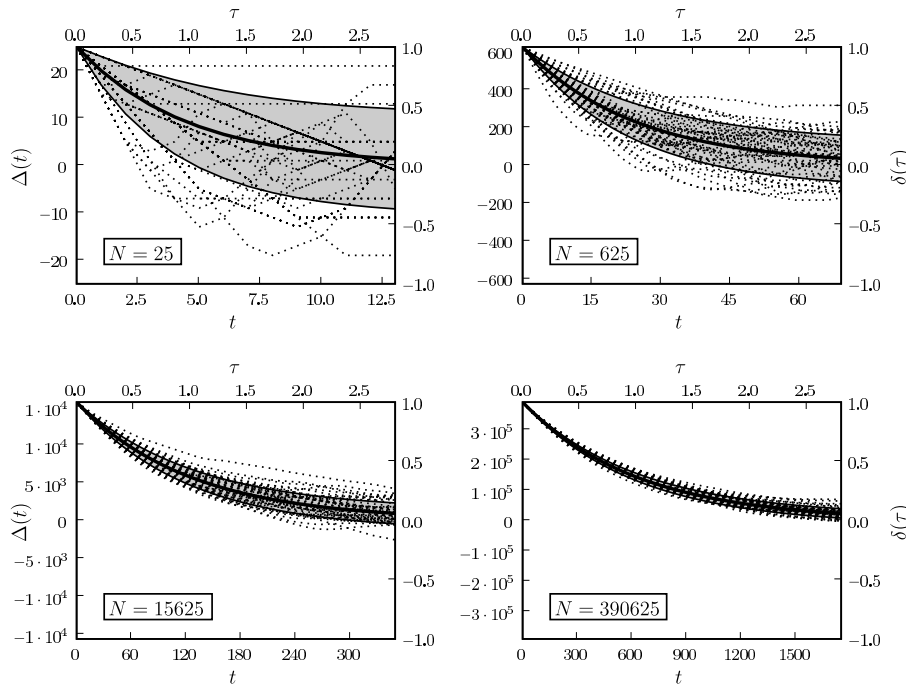


Fig. 6 Convergence of trajectories to the ensemble average as the size of the Kac ring grows large. The two axes are labeled both in unscaled and scaled variables with $\alpha = \beta = \frac{1}{2}$; the viewpoint is chosen such that the range of the graphs is identical with respect to the scaled limit variables.

The answer to this question is, perhaps surprisingly, *no!* It is just another single ordering of cards out of a total of $N!$ possible such arrangements. In other words, it is one single outcome from the sample space of size $N!$ corresponding to the experiment of randomly permuting N cards. Thus, to learn anything interesting, we must reconsider the question we ask.

A first possible direction is to ask about the shuffling itself. Does the process reduce *correlations* between the initial and the final states? This is an important topic, but not one we want to pursue here [1, 16]. A second direction is the study of *coarse-grained* or *macroscopic* descriptions. Here, a number of microscopic states is clustered by a many-to-one map into a much smaller number of *macroscopic states*. We can then talk about macroscopic states with relatively many microscopic realizations as being more disordered than those with relatively fewer realizations—the “disorder” of an event is related to its probability of occurrence in the experiment. Of particular interest is the limit of large system size where $N \rightarrow \infty$.

For our deck of cards, we can *define* a family of ordered macroscopic states as follows. For each $0 \leq n \leq N$, consider the macroscopic state that the first n cards of the standard ordering make up the top n cards of the stack and, consequently, the remaining $N - n$ cards of the standard ordering make up the bottom part of the stack. Such states are manifestly macroscopic as they can be realized via $n!(N - n)!$ different microscopic states. (The number of possible realizations is just the number of permutations that leave the top n -deck and the bottom $(N - n)$ -deck invariant.) Relative to the total number of available microstates (the total number of permutations of N

cards), this number is small. In other words, the probability

$$\begin{aligned} P(n) &= \frac{\text{number of microstates in the ordered macrostate}}{\text{total number of microstates}} \\ (36) \quad &= \frac{n!(N-n)!}{N!} = \binom{N}{n}^{-1} \end{aligned}$$

that a random shuffle yields an ordered state is small—the “dynamics” will prefer disordered states. For example, for a deck of $N = 8$ cards,

$$(37) \quad P(N/2) = \frac{4! \cdot 4!}{8!} = \frac{1}{70},$$

while for a standard deck of $N = 52$ cards,

$$(38) \quad P(N/2) = \frac{26! \cdot 26!}{52!} \approx 1.6 \cdot 10^{-16}.$$

More generally, as N becomes large, the probability of obtaining an ordered state from a random shuffle goes to zero. We also remark that among the ordered states according to the above definition, the state where $n = N/2$ is the most ordered, due to the following fact.

Exercise 11. Prove that for any $0 \leq k \leq n$,

$$(39) \quad \binom{n}{k} \leq \binom{n}{\lfloor n/2 \rfloor},$$

where $\lfloor x \rfloor$ denotes the largest integer less than or equal to x .

Exercise 12. Show that the half-width about the maximum of the binomial coefficient function $k \mapsto \binom{n}{k}$ behaves asymptotically like $\sqrt{2n \ln 2}$ as $n \rightarrow \infty$. (This result is ultimately connected to the approximation of a binomial distribution with large n and small skew by a normal distribution, a fact well known in probability theory.)

The card-shuffling example provides a good illustration of coarse-graining. In particular, it shows that the quantitatively relevant concept is the number of microstates available to a given macrostate, while the notions of “order” and “disorder” are at best incidental. On the other hand, the example lacks a macroscopic quantity which is easily scaled. (Its state space is the group of permutations of the deck.) Let us therefore return to the Kac ring as it is initially set up. It has two important features.

1. The system is made up of N independent identical components which can be in one of two possible states.
2. The macroscopic observable is proportional to the number of components in each state.

These features generalize to many real physical systems.

We first introduce the *partition function* Ω as the number of microstates for a given macrostate. The macrostate is fully specified by Δ or, equivalently, by the number of black balls $B = (N + \Delta)/2$ or the number of white balls $W = N - B = (N - \Delta)/2$. The state with B black balls and $W = N - B$ white balls can be realized in

$$(40) \quad \Omega(B) = \binom{N}{B} = \frac{N!}{B!W!}$$

different ways.

The logarithm of this quantity, however, turns out to be more useful because it scales approximately linearly with system size when N is large, as we shall show below. This motivates the definition of the *Boltzmann entropy*

$$(41) \quad S = \ln \Omega .$$

Both S and Ω are functions of the macrostate Δ only.

Let $p = B/N$ denote the probability that a site carries a black ball and $q = W/N$ the probability that a site carries a white ball. Clearly, $p + q = 1$. We can then compute the behavior of the entropy for large N by using Stirling's formula in the form

$$(42) \quad \ln k! \sim k(\ln k - 1) \sim k \ln k$$

as $N \rightarrow \infty$. Hence,

$$(43) \quad S = \ln \Omega = \ln \frac{N!}{(pN)!(qN)!} \sim -N(p \ln p + q \ln q) .$$

In the continuum limit, we let, as before, $\delta = \Delta/N$ denote the macroscopic grayness, so that $p - q = \delta$. Since $p + q = 1$, both probabilities are functions of the macrostate δ only. Thus, the right-hand expression in (43) is a product of two terms, the first of which depends only on the size of the system, and the second only on the macroscopic state. In thermodynamics, such variables are called *extensive*, in contrast to *intensive* variables which do not depend on the system size. For example, mass and volume are extensive variables, while density and pressure are intensive. In our case, the entropy S is extensive while the grayness δ is intensive. We see that, in particular, the entropy is additive—if two macroscopic rings with the same grayness are joined, the entropy of the resulting ring is the sum of the component entropies.

We also note that the macrostate of maximal entropy is the *equilibrium state* $\delta = 0$ of the macroscopic Kac ring dynamics under the Stoßzahlansatz; see Exercise 13.

Remark 3. To define a continuum entropy in the sense of section 6, we may divide out the system size from (43), setting

$$(44) \quad H = \frac{S}{N \ln 2} = -(p \log_2 p + q \log_2 q) .$$

This expression is called *information entropy* [14] and gives the (relative) minimal number of bits of information, on average in the limit of large system size, required to fully specify the microstate of the Kac ring for a known macrostate Δ . The information entropy, however, is not extensive. To define a macroscopic extensive entropy, we may set $N = \rho V$, where ρ is the “particle density” and V the “volume” of the ring. Then the continuum limit can be written as $\rho \rightarrow \infty$ with V fixed, and

$$(45) \quad s \equiv S/\rho = -V(p \ln p + q \ln q)$$

is the resulting extensive continuum entropy.

Remark 4. Formally, we may think of δ and s as thermodynamic temperature and entropy, respectively. However, the Kac ring model is too simple to allow for any interesting thermodynamic behavior— s is trivially related to δ and there are no additional independent variables, so that thermodynamic cycles such as the steam engine cycle cannot be found within the model. We must therefore caution against reading too much into such formal analogies.

Exercise 13. Show that S in (43) is maximal for equiprobability, i.e., when $p = q = \frac{1}{2}$.

Exercise 14. Verify that the interpretation of information entropy given in Remark 3 makes sense for the neutrally gray state $\delta = \frac{1}{2}$ and for the uniformly black state $\delta = 1$.

Remark 5. In the case of a molecular gas, the microscopic variables, i.e., the positions and velocities of the molecules, take values in a continuum. Therefore, the number of microstates per macrostate has to be replaced by a state density function depending on the macroscopic variables temperature and pressure.

8. Ehrenfests' Urn Model. Paul (1880–1933, a student of Boltzmann) and Tatyana (born Afanasjeva, 1876–1964) Ehrenfest came up with the first simplified model to illustrate how to reconcile thermodynamics and irreversibility with the underlying reversible laws of classical mechanics [5].

N balls, labeled $1, \dots, N$, are distributed between two urns. Each time step, a number k is drawn from among the labels $1, \dots, N$ at random and the k th ball is moved from its current urn into the other. Now consider the macroscopic quantity

$$(46) \quad D(t) = |N_{\text{I}}(t) - N_{\text{II}}(t)|,$$

the difference between the number of balls in urn I and the number of balls in urn II, as a function of time.

Exercise 15. In which sense is the microscopic behavior of the Ehrenfest model time-reversible? Does it have recurrence?

Exercise 16. Define an entropy function for this system. What are the states of maximal and minimal entropy?

Exercise 17. Find an equation for $\langle D(t) \rangle$. How do $D(t)$ and $\langle D(t) \rangle$ compare on a fixed time interval as $N \rightarrow \infty$? For fixed N as $t \rightarrow \infty$?

Exercise 18. Is there a continuum limit as in section 6?

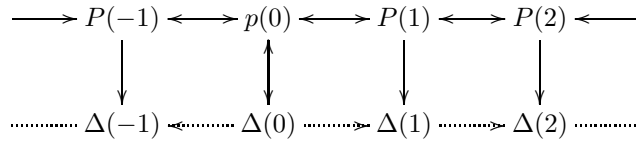
Exercise 19. Write a computer program to simulate the Ehrenfest model.

9. Discussion. In this paper we have exemplified the process of coarse-graining via the Kac ring model. Let us reconsider the results once more from a distance.

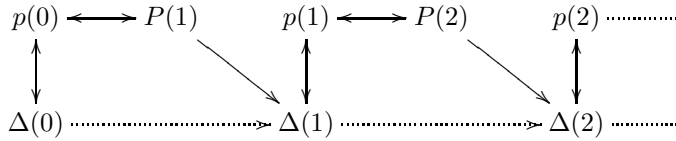
We are looking at a system on a very large microscopic state space. It can be decomposed into a large number of simple interacting subsystems—the balls and markers (or the molecules in a gas). The dynamical description at this level is deterministic and time-reversible. The microscopic state is assumed to be nonobservable; it cannot be measured or manipulated. We thus introduce a coarse-graining function, a many-to-one map from the microscopic state space into a much smaller macroscopic state space; the output of the coarse-graining function is the experimentally accessible quantity Δ . An evolution law for this coarse-grained quantity $\Delta(t)$ is therefore desirable.

Given that the coarse-graining function is highly noninvertible, we resort to a statistical description. For a given initial $\Delta(0)$, we construct an ensemble of systems such that the corresponding macrostate—the expected value of the coarse-graining function applied to the members of the ensemble—matches $\Delta(0)$. In the absence of further information, we must assume that all constituent subsystems are statistically independent. We can then describe the evolution of Δ in two different ways.

In the so-called *Liouville approach*, we evolve each member of the ensemble of microstates up to some final time T , apply the coarse-graining function to each member of the ensemble, and finally compute the statistical moments of the resulting distribution of macrostates (see Figure 7a). Clearly, the macroscopic evolution must now also



(a) Relation of microscopic to macroscopic quantities in the Liouville approach.



(b) Relation of microscopic to macroscopic quantities in the Boltzmann approach.

Fig. 7 *Liouville vs. Boltzmann approach. At the upper, microscopic level, p and P denote probability functions for the ensemble. When the subsystems are statistically independent, the probability function is a product of single-subsystem probability functions, symbolically written p . When the subsystems lose their statistical independence, we need a probability function for the entire system, symbolically denoted P . The microscopic evolution destroys statistical independence. In the Liouville approach, even when starting with a distribution $p(0)$ of statistically independent subsystems at $t = 0$, we keep the full probability function $P(t)$ at every $t \neq 0$. Thus, it is only the application of the coarse-graining function which is irreversible when $t \neq 0$. In the Boltzmann approach, statistical independence is forcibly restored at every time step, thereby creating an “arrow of time.”*

be described in terms of the ensemble mean. So the best we can hope for is that the variance of the final distribution of macrostates is small. Then a single performance of the experiment will, with high probability, evolve close to the ensemble mean. Only then does coarse-graining make sense at all. For the Kac ring model, for example, we were able to prove in sections 4 and 5 that the variance remains relatively small over a sufficiently small number of time steps. In general, however, finding an explicit expression for the mean and variance is impossible. Moreover, the Liouville approach will often be intractable by numerical computation, too—even the simulation of a moderately large ensemble of Kac rings on a desktop computer takes a nonnegligible amount of time.

The second approach, the so-called *Boltzmann* or *master equation* approach, is based on the experience that, under the assumption of statistical independence of the subsystems, it is usually easy to predict the macroscopic mean after one time step. (Remember that the computations in the Boltzmann section 3 are much easier than those in the Liouville section 4!) However, the interaction of subsystems during a first time step will generally damage their statistical independence, if only slightly. Still, we might pretend that, after each time step, all subsystems are still statistically independent. In other words, we might choose to keep only macroscopic information across consecutive time steps, forgetting all other information about the microstates. This approximation is nothing but the *Stoßzahlansatz*. In general, the resulting macroscopic dynamics will differ from the Liouville dynamics after more than one time step. (For the Kac ring, they differ only when $t > N$, but this is rather exceptional.)

From the description above, it is clear that both approaches to coarse-graining break the time-reversal symmetry of the microscopic dynamics, as the coarse-graining map from an ensemble of microstates to the macroscopic ensemble mean is invertible

if and only if the constituent subsystems are statistically independent. Thus, the loss of statistical independence defines a macroscopic arrow of time (see Figure 7b). Thus, with coarse-graining interpreted this way, the Loschmidt paradox does not exist.

It is worth emphasizing that this point of view is the correct interpretation pertaining to real-world experiments whenever specific microstates cannot be prepared. Any such experiment will, with probability approaching one as the system size tends to infinity, start out in a typical microstate for the given macrostate. This does not exclude the existence of nontypical initial states. In particular, when we assume that the final state of the experiment is taken as the initial state of the time-reversed experiment, we are clearly considering nontypical initial data. If this were experimentally possible, it would mean that there exists a macroscopic way to correlate the individual subsystems *before* they interact with each other. When such control is part of the experiment, it must become part of the analysis as well, requiring a refinement of the macroscopic state space with subsequent changes to the notion of entropy.

So how does Zermelo's paradox fare? Again, no surprise. The Stoßzahlansatz in the Boltzmann approach is an approximation which depends on weak statistical dependence of the subsystems. However, as time passes, interactions will increase statistical dependence. Thus, we cannot expect that the approximate macroscopic mean remains a faithful representation of the recurrent microscopic dynamics over a long period of time. On the contrary, it is the validity of the Stoßzahlansatz over any time scale which requires justification. Figure 6 gives a sense that for the Kac ring, the time scale of relevance of the approximate dynamics is very much shorter than the recurrence time.

In real physical many-body systems, the times of (approximate) recurrence are astronomically large compared to typical time scales of thermodynamic experiments. Moreover, the chaotic character of the evolution will help maintain statistical independence over long intervals of time. This is illustrated by a computer simulation described in [3]. A gas is prepared in one side of a box which is separated into two compartments only accessible via a small hole. During the simulation, the molecules pass through the small hole and the gas, slowly, distributes equally between both halves of the box toward statistical equilibrium. If the velocities are now reversed, the gas will indeed return back to its initial condition with all molecules confined to one compartment. However, this time-reversed state is unstable: if it is slightly perturbed, the system will remain in equilibrium. The fluctuations which decorrelate the prepared time-reversed state can be very small, but will be amplified through the chaotic nature of many-particle dynamics. This also explains why, in most cases, it is infeasible to design "hidden controls" to create statistical dependence between molecules in the initial state of a thermodynamic experiment.

Exercise 20. For given $\Delta > 0$, set up a Kac ring in its initial state S_0 with a random distribution of markers and balls. First, turn the ring by one clockwise step into a new state S_1 . Is Δ more likely to increase or to decrease during this first step? Next, turn the ring one step counterclockwise. Since the dynamics is time-reversible, we are reverting back to S_0 . Finally, turn the ring yet another step counterclockwise into a state S_{-1} . Is Δ now more likely to increase or to decrease?

10. Classroom Notes. Students of mathematics and physics usually have to endure a long period of acquiring methods and techniques and of learning the necessary formalism before they can embark on tackling deep and fundamental problems that widen their intellectual horizons, not unlike the experience of learning a foreign language, where first one has to learn a substantial vocabulary before meeting and

conversing with friends. However, very often these fundamental questions were the very reason students chose mathematics or physics as their subject in the first place.

The lectures on which this paper is based try to engage students in very fundamental and deep concepts with only a very basic knowledge of mathematics. They were developed when we were asked, independent of one another at our respective institutions, to teach a course for first-year students that is orthogonal to the standard fare of calculus and linear algebra, but rather stimulates the students' interest, creative thinking, and exposure to various branches of mathematics.

Visits to the library soon revealed a wealth of elementary but beautiful and deep problems paradigmatic for many areas of pure mathematics. The supply on the applied side, however, is much more scarce. Where are the problems that require only elementary tools without being isomorphic to some page of a reform-calculus textbook, that are applied without being narrow, or that involve theory as well as computation or experiment? The obligatory scan of the Education section in *SIAM Review* over the years also yields a wealth of stimulating ideas. But for first years, yet still with some sense of broad relevance, there are basically no resources.

Thus, we were thrilled when one of us (G.G.) discovered the Kac ring in one of the introductory chapters of Dorfman's book on nonequilibrium statistical mechanics [3]. We have now used this topic as a two- or three-week unit within our courses several times. Our set of notes has subsequently grown from an extract of Dorfman to something more self-contained and, we hope, independently useful. While we might be pushing the notion of "for first year students" at times, we have written this paper very much with first years in mind and believe that an experienced instructor will find it easy to distill the material down to an appropriate level of detail in a setting similar to ours.

In our experience, the topics covered in this paper can easily be taught in about 4–6 lecture hours, depending on the depth of coverage and mathematical maturity of the students. We believe that the material offers two large pedagogical benefits.

First, very little prior hard knowledge is necessary. The concepts from combinatorics and probability are so elementary that they can be taught "on the fly," if necessary, or the lectures may be embedded into a more general first introduction to combinatorics and probability. For the most part, the discussion does not need calculus beyond elementary limits which students will encounter early on in their regular first year calculus class; only the discussion of the continuum limit in section 6 requires a deeper understanding of limits, but may be safely omitted or explored through computation. Section 5 is conceptionally important and ultimately not more difficult than the derivation of ensemble averages in the previous section, but the algebra can easily be intimidating and may be deemphasized.

Second, the Kac ring (or the Ehrenfest model, which makes a good supplementary assignment) can easily be explored computationally with few lines of code in basically any programming language on any platform. The transition from microscopic to macroscopic behavior of the Kac ring can be visualized by arranging the sites on concentric rings, the outermost representing the microscopic dynamics, each subsequent smaller ring being obtained through averaging of two neighboring sites, with the macroscopic grayness value ultimately emerging at the center; see Figure 8. Python source code for the animation and all of the figures in this paper is available from the authors' websites.

Acknowledgments. We thank Vadim Kaimanovich and David Levermore for comments on earlier versions of this paper, and our students for finding many typos



Fig. 8 Snapshot of the Kac ring animation, the markers being represented by the outermost black-and-white ring, and the grayness color-coded such that red represents $\delta = 1$, blue represents $\delta = -1$, and white the equilibrium where $\delta = 0$. The different rings show successive coarse-graining via next-neighbor averaging in a binary tree fashion.

and useful suggestions. Any remaining errors and misconceptions are necessarily ours. The code for the Kac ring animation (see Figure 8) was cowritten by Jacobs student Vitalie Patrinea.

REFERENCES

- [1] D. ALDOUS AND P. DIACONIS, *Shuffling cards and stopping times*, Amer. Math. Monthly, 93 (1986), pp. 333–348.
- [2] C. CERCIGNANI AND E. GABETTA, EDs., *Transport Phenomena and Kinetic Theory: Applications to Gases, Semiconductors, Photons, and Biological Systems*, Birkhäuser, Boston, 2007.
- [3] J.R. DORFMAN, *An Introduction to Chaos in Nonequilibrium Statistical Dynamics*, Cambridge University Press, Cambridge, UK, 1998.
- [4] M. DRESDEN, *New perspectives on Kac ring models*, J. Statist. Phys., 46 (1987), pp. 829–842.
- [5] P. EHRENFEST AND T. EHRENFEST, *Über zwei bekannte Einwände gegen das Boltzmannsche H-Theorem*, Phys. Z., 8 (1907), pp. 311–314.
- [6] M. KAC, *Some remarks on the use of probability in classical statistical mechanics*, Acad. Roy. Belg. Bull. Cl. Sci. (5), 42 (1956), pp. 356–361.
- [7] J.L. LEBOWITZ, *Macroscopic laws, microscopic dynamics, time's arrow and Boltzmann's entropy*, Phys. A, 194 (1993), pp. 1–27.
- [8] D.J. PINE, J.P. GOLLUB, J.F. BRADY, AND A.M. LESHANSKY, *Chaos and threshold for irreversibility in sheared suspensions*, Nature, 438 (2005), pp. 997–1000.
- [9] H. POINCARÉ, *Sur le problème des trois corps et les équations de la dynamique*, Acta. Math., 13 (1890), pp. 1–270.
- [10] I. PRIGOGINE AND I. STENGERS, *Order Out of Chaos: Man's New Dialogue with Nature*, Bantam Books, Toronto, 1984.
- [11] J. ROTHSTEIN, *Loschmidt's and Zermelo's paradoxes do not exist*, Found. Phys., 4 (1974), pp. 83–89.
- [12] L.S. SCHULMAN, *Time's Arrow and Quantum Measurement*, Cambridge University Press, Cambridge, UK, 1997.
- [13] E. SEGRÉ, *From Falling Bodies to Radio Waves*, W. H. Freeman, New York, 1984.

- [14] C.E. SHANNON, *A mathematical theory of communication*, The Bell System Technical J., 27 (1948), pp. 379–423 and pp. 623–656. Available online at <http://cm.bell-labs.com/cm/ms/what/shannonday/shannon1948.pdf>.
- [15] T. SHINBROT, *Drat such custard!*, Nature, 438 (2005), pp. 922–923.
- [16] L.N. TREFETHEN AND L.M. TREFETHEN, *How many shuffles to randomize a deck of cards?*, R. Soc. Lond. Proc. Ser. A Math. Phys. Eng. Sci., 456 (2000), pp. 2561–2568.
- [17] J. UFFINK, *Boltzmann's Work in Statistical Physics*, in The Stanford Encyclopedia of Philosophy (Winter 2004 Edition), E.N. Zalta, ed. Available online at <http://plato.stanford.edu/archives/win2004/entries/statphys-Boltzmann/>.
- [18] E. ZERMELO, *Über mechanische Erklärungen irreversibler Vorgänge. Eine Antwort auf Hrn. Boltzmann's "Entgegnung,"* Wiedemann Ann., 59 (1896), pp. 793–801.