

# TIME-RELATED ISSUES IN STATISTICAL MECHANICS

L. S. Schulman

*Physics Department, Clarkson University,  
Potsdam, New York 13699-5820, USA\**

(Dated: November-December 2005 and March-April 2006)

Current version: April 30, 2008

Topics:

- The description of apparent of irreversibility
- Physical origins of the arrow(s) of time
- Two-time boundary value problems
- The micro/macro distinction and coarse graining
- Quantum mechanics with “special states.”

This version of the notes (issued in September 2006) contains an additional section (X E) on the implications of the last of the above topics for determinism.

*These lectures are part of Cours de physique théorique du SPhT, Année 2005-2006, as well as Physics-118112, Technion, Spring 2006.*

---

\*Electronic address: [schulman@clarkson.edu](mailto:schulman@clarkson.edu)

## Contents

INTRODUCTION	4
<b>I. Irreversibility</b>	4
A. The Kac ring model	5
1. The paradoxes	7
B. Resolution	7
C. Does this explain the arrow of time?	8
D. Information as $-\sum p \log p$	9
E. Classical and quantum recurrence theorems	10
1. The classical theorem. From [1].	10
2. The quantum theorem. From [1].	11
F. The Cat Map	12
1. Criteria for equilibration	15
G. The cat map is mixing	16
<b>II. Quantum irreversibility</b>	18
<b>III. Stochastic dynamics: irreversibility from coarse graining</b>	23
<b>IV. Initial conditions vs. final conditions: a restatement of the second law</b>	26
1. Past/future reliability: <i>not</i> a restatement of the second law	28
<b>V. Arrows of time</b>	29
<b>VI. Correlating arrows of time</b>	30
<b>VII. The universe with two-time boundary conditions</b>	33
1. How the universe imposes its arrow.	33
A. Solving two-time boundary value problems	34
B. Symmetric behavior of a non-time-reversal invariant dynamical map	38
<b>Two-time boundary conditions: Counting states in a cubic centimeter of air</b>	39
<b>VIII. How do you define macroscopic? ... with practical consequences</b>	41
A. The role of time	42
B. The framework: stochastic dynamics	42
C. Phase transitions	43
A basic relation	43
D. Many phases, extension of the basic relation	44
E. Concentration of the probabilities on the phases	47
F. Basins of attraction	48
G. Probability to arrive in a given phase—barycentric coordinates	48
H. Hierarchical metastable states	49
<b>IX. Coarse grains</b>	51
Random walk: time recovers space	52

A. Making grains	53
1. Dynamic Ferromagnet	53
2. Words (organization of concepts) and the brain	54
3. The karate club	55
B. Reflections	56
<b>X. Quantum measurement theory</b>	56
A. Precis of the idea: the schizoid cat	57
B. Existence of special states	59
A schizoid cloud chamber	59
Defining the Hamiltonian	60
Qualitative description of the operation of the detector	62
1. Spontaneous spin flip: false positives and ground state stability	62
2. Scattering-induced spin flips	63
3. Operation of the detector	64
4. Scattering-induced spin flips with SPECIAL initial conditions	65
5. Other apparatus models	66
C. Recovering the usual probabilities	66
$ \psi ^2$ as probability	68
D. Experimental tests	69
E. Excerpt from Ref. [1] ( <i>Time's Arrows and Quantum Measurement</i> ), Sec. 6.3, on the determinism implicit in this worldview.	70
<b>Acknowledgments</b>	70
<b>References and endnotes</b>	71

## INTRODUCTION

The overall reference for most of these lectures is my book [1], *Time's Arrows and Quantum Measurement*. Short excerpts from [1] appearing in these notes are used with permission of Cambridge University Press.

The focus of this lecture series is time asymmetry, the “arrows of time.” First we’ll talk about the thermodynamic arrow: how things can behave irreversibly, despite time symmetry in the underlying dynamics. A good deal of effort will be spent developing models and tools for this analysis.

Then we’ll examine another arrow, cosmological expansion, and discuss the hypothesis that the cosmological and thermodynamic arrows are related. This will introduce a more time-symmetric way of looking at things, and we’ll talk about “two-time boundary conditions.” Having related thermodynamics and cosmology, there arises the possibility that there are thermodynamics-based approaches to deducing the long-term future of our universe.

Then a digression on the definition of entropy. If the second law (of thermodynamics) is fundamental, how can it depend on subjective definitions? The subjectivity I refer to arises in the distinction between “work” and “heat,” or between macroscopic and microscopic. Here I’ll use methods of non-equilibrium statistical mechanics, and (perhaps) surprisingly this will turn out to have implications for a practical contemporary issue: the identification of communities in a network.

Finally, we’ll take up quantum measurement theory. I’ll tell you about a suggestion I’ve made for retaining unitary quantum time evolution throughout the “measurement” process and nevertheless avoiding superpositions of macroscopically different states. Although this resolution of the measurement problem tampers not at all with quantum mechanics, it plays havoc with the foundations of statistical mechanics. We are used to the idea that to calculate the future of a macroscopically defined system one assumes that all compatible microstates are present with equal probability. It is this assumption that is dropped, and I am hoping that the perspective you will have gained from the “time’s arrows” part of these lectures will make this suspension less painful. Experimental tests of these ideas will also be discussed.

### I. IRREVERSIBILITY

The problem is the apparent contradiction between microscopic reversibility and the macroscopic arrow. (Later I’ll say why I don’t think CP violation has any bearing on these issues.)

A precise display of this contradiction comes from the Boltzmann H-Theorem. The theorem is based on the equation of motion of the quantity  $f(\mathbf{r}, \mathbf{v}, t)$ , which is the number of particles, at time  $t$ , in the volume  $d^3\mathbf{r}d^3\mathbf{v}$  around  $(\mathbf{r}, \mathbf{v})$ , with  $\mathbf{r}$  a point in coordinate space and  $\mathbf{v}$  a point in velocity space. Boltzmann found that  $f$  satisfies the transport equation

$$\left( \frac{\partial}{\partial t} + \mathbf{v}_1 \cdot \nabla_{\mathbf{r}} + \frac{\mathbf{F}}{m} \cdot \nabla_{\mathbf{v}_1} \right) f_1 = \int d\Omega \int d^3v_2 \sigma(\Omega) |\mathbf{v}_1 - \mathbf{v}_2| (f_2' f_1' - f_2 f_1). \quad (1)$$

The subscripts on  $f$  refer to its arguments. The left hand side describes the “flow” of  $f$  due to the velocity of the particles and to any external forces, where  $\mathbf{F}$  is an external force,  $m$  the molecular mass, and  $f_1 \equiv f(\mathbf{r}, \mathbf{v}_1, t)$ . On the right is the change in  $f$  due to scattering. As for  $f_1$ , the subscripts on the  $f$ ’s in the integral are the subscripts on its argument  $\mathbf{v}$ . The

quantity  $\sigma(\Omega)$  is the center-of-mass scattering cross section for a pair of molecules whose relative velocity goes from  $\mathbf{v}_2 - \mathbf{v}_1$  to  $\mathbf{v}'_2 - \mathbf{v}'_1$ . The scattering term is derived by assuming that the scattering of particles 1 and 2 is independent of their previous history. This is the assumption of molecular chaos, the *Stosszahlansatz*. Let  $\mathbf{F} = 0$ . Then we can take  $f$  independent of  $\mathbf{r}$ . Define

$$H(t) \equiv \int d^3v f(\mathbf{v}, t) \log f(\mathbf{v}, t). \quad (2)$$

This is essentially the negative of an entropy. An immediate consequence of Eq. (1) is that  $\frac{dH}{dt} \leq 0$ . This is the H-theorem and it establishes that entropy never decreases, even though the dynamics of the system is symmetric in time. (For a derivation of Eq. (1) and the proof of the last assertion, see Huang [2]. Why  $H$  is a kind of entropy will be discussed in Sec. ID below.)

There are two famous problems connected with this result. There is the *reversibility* paradox, associated with the name Loschmidt, and which is basically the conundrum of deriving something asymmetric from symmetric assumptions. It can be phrased as follows: Let the system go forward a time  $t$  and suppose that  $H$  has actually decreased. Now reverse the velocity of every particle. On the one hand, this system has the same dynamics as before, just different initial conditions. So  $H$  should continue to decrease (or at least not increase). But the reversal of velocities causes the system to retrace its path, returning to a higher value of  $H$  (its initial condition).

The other problem is the *recurrence* paradox, based on a theorem of Poincaré. For a system in a finite volume having a potential that is not too wild, the surface of constant energy in phase space has finite volume. Under these conditions almost all initial states eventually return (arbitrarily closely) to themselves. This means that although for a while  $H$  may decrease, eventually it must increase.

### A. The Kac ring model

Marc Kac invented a lovely model suitable for demonstrating the above paradoxes. He also showed what it is you really *can* prove. To wit, although one particular system is not guaranteed to satisfy the H-theorem, for *ensembles* of systems (and with appropriate caveats) you really can prove the decrease of  $H$ , or increase of entropy.

We will work through this model in detail.

Consider a ring with  $N$  sites on it. See Fig. 1, where the sites are pictured as little cups. In every site there is a ball, which can be either black or white. On each time step, every ball moves counterclockwise to the next site. A subset of the sites are designated “active.” Let there be  $A$  of these “active” sites (the symbol  $A$  will also designate the set) and let them have the following property: when a ball *leaves* an active site, it switches color, turning white if it was black, black if it was white. In the figure  $N = 25$  and  $A = 5$ .

Let the number of white balls in active sites at time  $t$  be  $W_A(t)$ , and let  $B_A(t)$  be correspondingly defined. The dynamical scheme implies

$$W(t+1) = W(t) - W_A(t) + B_A(t) \quad (3)$$

$$B(t+1) = B(t) - B_A(t) + W_A(t). \quad (4)$$

This is exact. Define  $\mu \equiv A/N$ , the active site fraction. For randomly distributed active

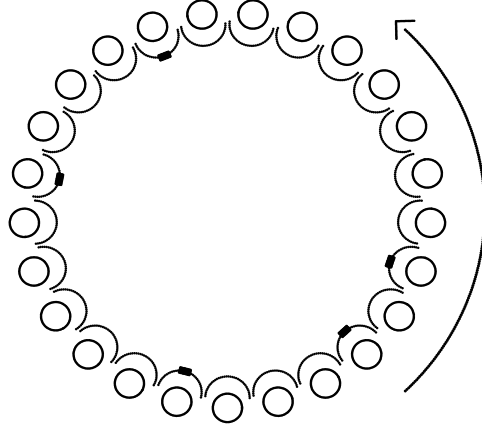


FIG. 1: Kac ring model. Balls move counterclockwise from site to site. Some sites are active and are marked at their bases. When a ball leaves an active site it changes color. Ball color is not shown.

sites it is reasonable to assume that

$$\frac{\text{number of white balls in active sites}}{\text{number of white balls everywhere}} = \frac{\text{number of active sites}}{\text{number of all sites}} \quad (5)$$

Thus  $W_A(t)/W(t) = \mu$  and similarly for  $B$ . (This is the analog of Boltzmann's molecular chaos assumption.) With the relation (5), Eq. (4) becomes

$$W(t+1) = W(t) - \mu[W(t) - B(t)] \quad (6)$$

$$B(t+1) = B(t) - \mu[B(t) - W(t)] \quad (7)$$

Subtracting the second equation from the first, the normalized difference on time step  $t+1$  is found to be

$$\delta(t+1) = \frac{W(t+1) - B(t+1)}{N} = (1 - 2\mu)\delta(t)$$

It follows that  $\delta(t) = (1 - 2\mu)^t \delta(0)$ . The system goes gray exponentially rapidly, with oscillation if  $\mu > 1/2$ .

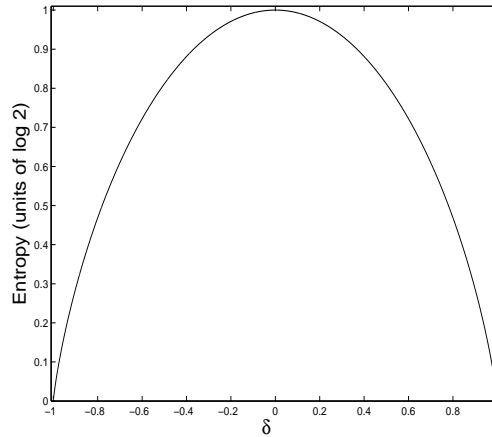
To discuss the paradoxes I'll define an entropy for this system and show that it is monotonically increasing. The general principle for an entropy calculation (as we'll momentarily establish in Sec. I D) is that it is the logarithm of the number of microscopic states consistent with the given macroscopic state (N.B. the i/a, micro/macro, distinction!). This leads to two problems. One is traditional: counting. The other is seldom mentioned: how do you define "macroscopic"? Later we will go into this problem, but for now we will take the quantity just defined,  $\delta$ , as the macroscopic variable.

Entropy is thus given by

$$\frac{S(\delta)}{k} = \log[\#\text{microstates consistent with } \delta \text{ (or } W)] = \log \left( \frac{N}{W} \right), \quad (8)$$

where  $k$  is a constant related to the base of the logarithms, and often, for historical (and occasionally substantive [3]) reasons, given physical units. Using Stirling's approximation, this yields

$$\frac{S}{kN} = -\frac{1+\delta}{2} \log \frac{1+\delta}{2} - \frac{1-\delta}{2} \log \frac{1-\delta}{2}, \quad (9)$$

FIG. 2: Entropy as a function of  $\delta$ .

which should be familiar as the information entropy (the amount of information missing, per ball, if all you know is  $\delta$ ). Cases: all white,  $S = 0$ ; all gray,  $S = kN \log 2$ .

To get a version of the H-theorem it is not necessary to differentiate  $S$ . The form of  $S$  as a function of  $\delta$  is shown in Fig. 2. The maximum is at  $\delta = 0$ . Moreover,  $|\delta|$  is a monotonically decreasing function of  $t$ . It follows that  $S$  is monotonically increasing. This is our “H-theorem.”

### 1. The paradoxes

The reversibility paradox makes use of a process very similar to the one just described: let the balls go *clockwise* and change color when *entering* a site. We could go through the arguments given previously and would again find that  $\delta(t)$  decreases in magnitude by the same factor,  $|1 - 2\mu|$ . It follows that if we start, say, with all white balls and take  $T$  steps with the first process, followed by  $T$  steps with the second process, we should have  $\delta = (1 - 2\mu)^{2T}$ . However, the second process is precisely the inverse of the first. Therefore in fact  $\delta$  is again 1.

The recurrence paradox arises from the simple observation that after  $2N$  time steps every ball has been in every site exactly 2 times. It has therefore changed color an even number of times and the original configuration has been restored. Therefore  $\delta$  *cannot* have been monotonically decreasing.

## B. Resolution

Make the proof more formal to see what’s happening:

Define

$$\epsilon_p \equiv \begin{cases} -1 & \text{if } p \in A \\ 1 & \text{if } p \notin A \end{cases}, \quad p = 1, \dots, N \quad \text{site variables} \quad (10)$$

$$\eta_p \equiv \begin{cases} +1 & \text{if ball-}p \text{ is white} \\ -1 & \text{if ball-}p \text{ is black} \end{cases}, \quad p = 1, \dots, N \quad \text{ball variables} \quad (11)$$

Then the equation of motion is

$$\eta_p(t) = \epsilon_{p-1} \eta_{p-1}(t-1), \quad (12)$$

which implies

$$\eta_p(t) = \epsilon_{p-1} \epsilon_{p-2} \dots \epsilon_{p-t} \eta_{p-t}(0). \quad (13)$$

For  $\delta$  we have (with some index shifts)

$$\delta(t) = \frac{1}{N} \sum_p \eta_p(t) = \frac{1}{N} \sum_p \prod_{j=0}^{t-1} \epsilon_{p+j} \eta_p(0). \quad (14)$$

The big step in resolving the paradoxes is to give up the idea of proving entropy increase for *individual systems*, and instead only try to establish the result for an ensemble of like systems. This is the cure prescribed in modern statistical mechanics. Let's see how it plays out for the ring model.

Usually one takes ensembles of initial conditions, but here we'll take an ensemble of active-site locations. That is, we'll consider many rings on each of which the specific subset that constitutes  $A$  is a random variable. In his original notes [4] and in his book [5], Kac fixes the exact number of sites in  $A$  and averages over their distribution. This is a worthwhile exercise in statistical mechanics, but we'll take an easier path. We'll simply flip a biased coin for each site and with probability  $\mu$  set it to be "active". (This is analogous to using the grand canonical ensemble rather than the canonical ensemble.) The  $\epsilon$ 's are thus independent identically distributed random variables ("i.i.d.") with expectation value

$$\langle \epsilon \rangle = (-1)\mu + (+1)(1 - \mu). \quad (15)$$

Then

$$\langle \delta(t) \rangle = \frac{1}{N} \sum_p \left\langle \prod \epsilon_{p+\ell} \right\rangle \eta_p(0). \quad (16)$$

Since the coin flips are not correlated the expectation of the product is the product of the expectations, and

$$\langle \delta(t) \rangle = \frac{1}{N} \sum_p \langle \epsilon \rangle^t \eta_p(0) = (1 - 2\mu)^t \delta(0). \quad (17)$$

This is a real proof, but what about the paradoxes?

Consider the recurrence paradox. After  $N$  steps, both  $\epsilon_p$  and  $\epsilon_{N+p}$  both appear in the product. They are the same site—and as far from being uncorrelated as you can get. So instead of  $\langle \epsilon_p \epsilon_{N+p} \rangle = (1 - 2\mu)^2$ , you have  $\langle \epsilon_p \epsilon_{N+p} \rangle = \langle (\pm 1)^2 \rangle = 1$ . Thus, in doing the proof you must be careful that the absence-of-correlation assumption is truly justified.

The reversibility paradox is handled similarly.

**Exercise:** Write down the "equation of motion" for the inverse process and show how correlations enter when contemplating forward and backward motions.

### C. Does this explain the arrow of time?

No, it does not. We've just seen in detail how a time-asymmetric result can be derived, but it only helps to sharpen the quandary over the source of our experiential arrow. On the



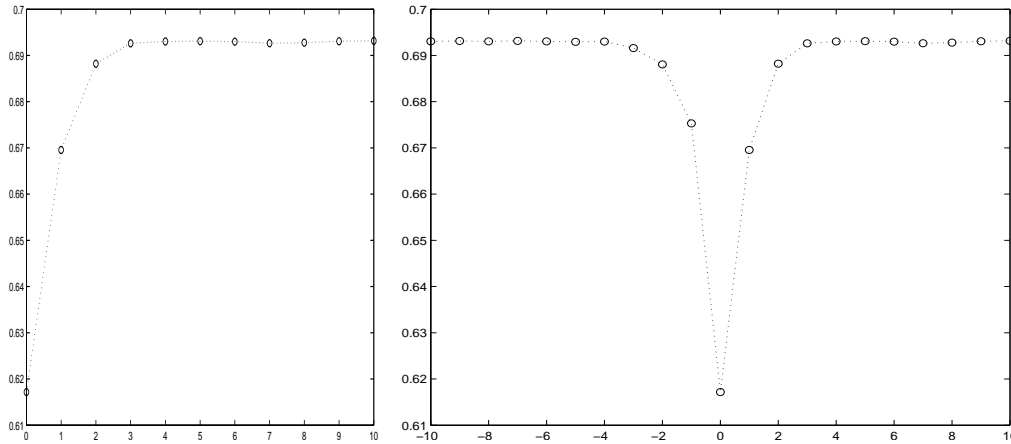


FIG. 3: Entropy as a function of time with initial conditions only, and for the same time-zero state working back to the state at earlier times.

left side of Fig. 3 is displayed the entropy increase for a system for which  $\delta(0) \approx 0.7$ . But if the same (time-0) ball-state is used to work back to negative times, entropy *also* increases, as is shown in the right-hand illustration in the same figure. This is exactly the conceptual point, in addition to the paradoxes, that one must confront in dealing with the Boltzmann H-theorem. (Recall that Boltzmann himself postulated an enormous fluctuation to account for the thermodynamic arrow [6].)

Why then does  $S$  increase away from zero? It's simple. We fixed a low entropy (i.e., large  $|\delta|$ ) state at time zero. This is a very unusual state, and when you do more or less anything to it, it becomes a *less* unusual state, i.e., its entropy increases. So the arrow arises only if you decide that time-0 is the beginning of your observations, which is something you usually decide because that's when you prepared your system. Prior to time-0 it was subject to other forces. So the source of the asymmetry is not the H-theorem, but is a consequence of your ability to prepare low entropy states as initial rather than final conditions.

Incidentally this view of the second law carries to other situations. Namely, if you start the system in an unusual state, and if it explores the state space, then it is more likely to be found in a less unusual state. This already implies an increase in entropy, making the second law something of a tautology. It's not quite a tautology, since the whole subject of ergodic theory concerns itself with whether or not the system does explore the entire state space. But if it does, you already have much of the content of the second law.

#### D. Information as $-\sum p \log p$

Here is a review of the connection between information and " $-\sum p \log p$ ." The reference is the book of Katz [7].

Suppose you have  $m$  boxes and there is a ball in one of them, but you don't know which. The information you are lacking is a function of  $m$ : If  $m$  is 1, you know everything; if  $m$  is large you know little. Call the missing information  $I(m)$ . Then  $I(1) = 0$  and  $I(m)$  is monotonically increasing with  $m$ .

Now suppose you have a rectangular array of boxes,  $m$  by  $n$ . The ball is in one of these  $mn$  boxes, and again you don't know which. So your missing information is  $I(mn)$ . On the

other hand, you could be directed to the correct box if someone would tell you the row and column in this rectangular array, so the missing information is also  $I(m) + I(n)$ . This leads to the fundamental relation

$$I(mn) = I(m) + I(n). \quad (18)$$

The solution to this functional equation can be found by assuming it holds for arbitrary real numbers, say  $I(xy) = I(x) + I(y)$ , for  $x, y \in \mathbb{R}^+$ . Then differentiate with respect to  $y$  and set  $y = 1$ . Call  $k \equiv I'(1)$ , which is nonnegative by the monotonic increasing property of  $I$ . Writing  $I'(x) = k/x$  leads to  $I(x) = k \log x$ .

Now we want to know the missing information if one has a probability distribution (knowing that the ball is in one of  $m$  boxes is also a probability distribution, but a trivial one;  $1/m$  for each box). Suppose there are  $n$  possibilities, with probabilities  $p_\alpha$ ,  $\alpha = 1, \dots, n$ . To deal with this we suppose the numbers  $p_\alpha$  are rational numbers, with common denominator  $N'$  (which may be very large). In fact, we even take a large multiple of  $N'$ , call it  $N$ , so that not only are the numbers  $Np_\alpha$  integers, but they are *large* integers. We now imagine  $N$  independent trials. Because  $N$  is so large, to a good approximation we can ignore fluctuations. Thus in  $Np_{\alpha_1}$  trials we'll get the result  $\alpha_1$ , in  $Np_{\alpha_2}$  trials we'll get the result  $\alpha_2$ , etc. What then is the missing information? We don't know in what order these were selected. So the missing information is the logarithm of the number of ways in which we can get the above outcome. This is

$$I_{N \text{ trials}} = k \log \frac{N!}{\prod_{\alpha=1}^n (Np_\alpha)!}. \quad (19)$$

By Stirling's approximation this is

$$I_{N \text{ trials}} = -Nk \sum_{\alpha=1}^n p_\alpha \log p_\alpha + O(\log N). \quad (20)$$

The missing entropy per trial is  $1/N$  times this:

$$I = -k \sum_{\alpha=1}^n p_\alpha \log p_\alpha, \quad (21)$$

a basic formula of information theory.

## E. Classical and quantum recurrence theorems

### 1. The classical theorem. From [1].

The Poincaré recurrence theorem is frugal in its assumptions, immediate in its proof and far reaching in its consequences. All you need is a measure-preserving transformation that is one-to-one and maps a set of positive, finite measure into itself under the action of the transformation. The cat map of Sec. IF satisfies this, and more importantly, so does the time evolution of a confined system in classical mechanics.

Because dynamical systems often possess small subsets for which atypical behavior can occur (e.g., for the cat map  $(0, 0)$  goes nowhere, while  $(0.4, 0.2)$  is periodic) it is necessary to formulate the recurrence theorem using the language of measure theory. You can think of Lebesgue measure as ordinary length, area or volume, generalized so that you can talk

about the measure of sets defined through limit processes. Countable sets or sets of lower dimension (lines in the plane, planes in 3-space) have measure zero. A useful notion once you have Lebesgue measure is “almost always” or “almost everywhere,” referring to statements that are true except possibly on a set of measure zero.

The theorem states that almost all points return arbitrarily closely to their initial positions. To state this more precisely, we make the following definitions:

For each  $t$ , a continuous or discrete real parameter, let a transformation  $\varphi_t$  map a space  $\Omega$  of finite measure into itself in a measure preserving way. That is, if  $\mu$  is the measure and  $A$  a subset of  $\Omega$ ,  $\mu(\varphi_t(A)) = \mu(A)$  for all  $t$ .

Let  $\varphi_t \varphi_s = \varphi_{t+s}$ .

We further assume:  $\varphi(\Omega) \subset \Omega$ ,  $\mu(\Omega) < \infty$ , and  $\varphi_t$  has an inverse.

It follows that if  $A \subset \Omega$ , then for almost all  $\omega \in A$ , there are arbitrarily large values of  $t$  such that  $\varphi_t(\omega) \in A$ .

To prove this, pick a time  $\tau$  and let  $B$  be the subset of  $A$  that does not recur after a time  $\tau$ , i.e., for all  $t \geq \tau$ ,  $\varphi_t(B) \cap A = \emptyset$ . We want to prove that  $B$  has measure zero. By definition, for all  $t \geq \tau$ ,  $\varphi_t(B) \cap B = \emptyset$ . Let  $\Phi \equiv \varphi_\tau$ ; it is sufficient to restrict discussion to powers of  $\Phi$ . The proof examines the orbit of  $B$ , i.e., the set of points  $\Phi^n(B)$  for  $n = 0, 1, \dots$ . By the hypothetical non-recurring property of  $B$  it follows (I’ll show this in a moment) that the orbit does not intersect itself, so all those images of  $B$  are disjoint. But if they are disjoint their measure adds and if the measure of  $B$  were anything but 0 the orbit would have infinite measure, which is impossible since that orbit is a subset of the finite-measure set,  $\Omega$ . The non-intersecting statement for the orbit is the following: for every  $n, m \geq 0$  and  $n \neq m$ ,  $\Phi^n(B) \cap \Phi^m(B) = \emptyset$ . To see this, suppose it to be false and that the intersection is a non-empty set  $C$ . Let  $n < m$ . Apply  $\Phi^{-n}$  to both sides to get  $B \cap \Phi^{m-n}(B) = \Phi^{-n}(C)$ . Because  $[(C \neq \emptyset) \Rightarrow (\Phi^{-n}(C) \neq \emptyset)]$ , this would give us a point in  $B$  that recurs after time  $(m - n)\tau$ , contrary to hypothesis.

## 2. The quantum theorem. From [1].

Consider a system confined to a box and let it have wave function  $\psi(0)$  at time 0. The quantum recurrence theorem states that the time-evolved wave function gets arbitrarily close to  $\psi(0)$  at arbitrarily large times. That is, for any  $\epsilon$  and any  $T_0$ , there exists a  $T$  such that  $\|\psi(T) - \psi(0)\| < \epsilon$ , with  $T > T_0$ .

As for the classical Poincaré recurrence theorem, the proof depends on the system’s being confined to a finite region. (A free particle just keeps going.) For the classical case, the finiteness was expressed by the requirement that the phase space volume “ $\Omega$ ” was finite. For quantum mechanics the discreteness of the spectrum does the job.

We will prove the theorem with a trick—appeal to the classical theorem [8]. The idea is that when you truncate the modes of a confined quantum system, its dynamics becomes that of a finite collection of classical oscillators. Then use the classical Poincaré recurrence theorem for those oscillators.

Let the given initial wave function be  $\psi(0)$ , call the Hamiltonian  $H$ , let its eigenvalues be  $E_n$  and the corresponding eigenfunctions  $u_n$ . Then at time  $t$  the wave function is given by

$$\psi(t) = \sum_n u_n r_n \exp(-iE_n t)$$

with  $r_n = \langle u_n | \psi(0) \rangle$ . Without loss of generality, we define the phases of the  $u_n$  so that the  $r_n$  are real. We must prove that for any  $\epsilon > 0$  and  $T_0$  we can find  $T > T_0$  such that

$$\|\psi(T) - \psi(0)\|^2 = 2 \sum_n r_n^2 (1 - \cos E_n t) < \epsilon$$

Since  $\sum_n r_n^2 = 1$ , we can find  $N$  such that

$$2 \sum_{n=N+1}^{\infty} r_n^2 (1 - \cos E_n t) \leq 4 \sum_{n=N+1}^{\infty} r_n^2 < \frac{\epsilon}{4}$$

The foregoing holds uniformly in  $t$ . It remains to show that there is a  $T > T_0$  such that

$$2 \sum_{n=1}^N r_n^2 (1 - \cos E_n T) < \frac{3\epsilon}{4} \quad (22)$$

This part of the proof is usually accomplished by quoting theorems about almost periodic functions. However, the sum in Eq. (22) can immediately be related to the motion of  $N$  harmonic oscillators—classical ones—and then we will be able to use the *classical* recurrence theorem. The idea is simple, but technicalities require a plague of  $\epsilon$ 's.

For the details of this plague consult [1].

The idea is that if you take  $N$  oscillators with frequencies  $E_n$ , then the sum in Eq. (22) compares their configuration at time-0 with that of a later time ( $T$ ). The classical recurrence theorem states that there will be times  $T$  (later than any  $T_0$  you care to give) for which the oscillators are all close to having  $\exp(iE_n T)$  be one. For these times the left-hand side of Eq. (22) is small and the theorem is proved.

**Remark:** Quantum recurrence allows no exceptions, not even as infrequent as the measure zero initial conditions of classical mechanics.

**Remark:** Note where the finite volume restriction played a role. It was essential to Eq. (22); discreteness of the spectrum allowed truncation to a finite sum.

**Remark:** It would be interesting to see whether things could be turned round and the Poincaré recurrence theorem used to establish results in the theory of almost periodic functions.

## F. The Cat Map

This is a richer dynamical system, a bit closer to classical mechanics. The universe—a mock phase space—is the unit square,  $\mathbb{I}^2$ . Points are designated  $(x, y)$ ,  $0 \leq x \leq 1$ ,  $0 \leq y \leq 1$ , and the transformation is

$$x' \equiv x + y \pmod{1} \quad (23)$$

$$y' \equiv x + 2y \pmod{1}. \quad (24)$$

Writing  $\xi = \begin{pmatrix} x \\ y \end{pmatrix}$ , the transformation can be alternatively expressed as

$$\xi' = \phi(\xi) \equiv M\xi \pmod{1}, \quad \text{with} \quad M = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}. \quad (25)$$

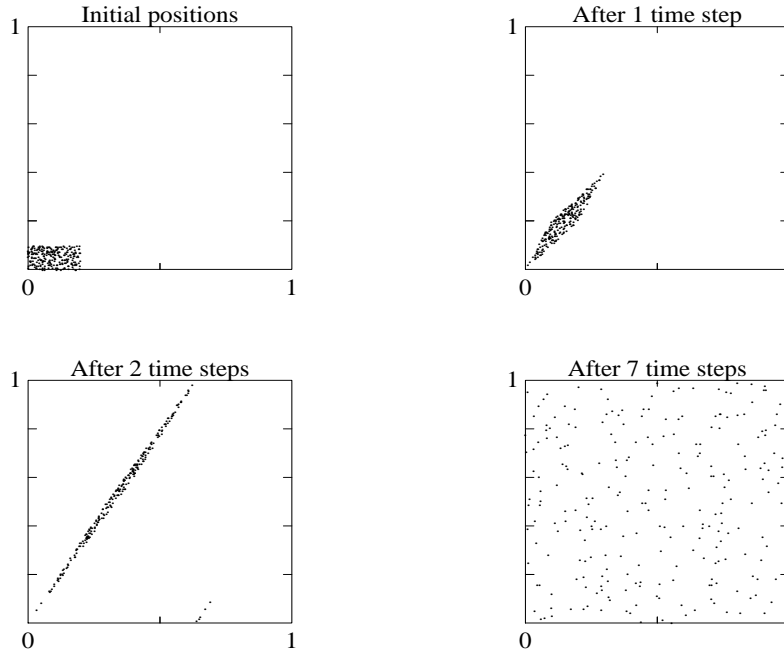


FIG. 4: Cat map for points initially near the origin.

The effect of this map on a number of points initially in a small rectangle near the origin is shown in Fig. 4. Evidently, if one thinks of these points as gas molecules, this system comes rapidly to equilibrium.

The reason for this equilibration can be seen in the spectrum of  $M$ . Its eigenvalues are

$$\lambda_{\pm} = \frac{3 \pm \sqrt{5}}{2} \approx e^{\pm 0.962} \approx \begin{cases} 2.618 \\ 0.382 \end{cases} \quad (26)$$

with eigenvectors [9]

$$v_{\pm} = \text{const} \cdot \begin{pmatrix} 1 \\ \lambda_{\pm} - 1 \end{pmatrix}. \quad (27)$$

By virtue of the large Lyapunov coefficient ( $\sim 0.96$ ) points along the direction  $v_+$  are stretched by a factor  $\sim 2.6$ , while those orthogonal to this direction (and along  $v_-$ ) are squeezed together (keeping constant area, since  $\det M = 1$ ). The action of  $\phi$  on a rectangle is shown in Fig. 5.

And finally, I would be remiss without an actual feline illustration, Fig. 6.

We wish a precise statement characterizing the apparent fact that this system comes to equilibrium. To do this, we consider a collection of points in  $\mathbb{I}^2$  to be an ideal gas of distinguishable particles and we coarse grain  $\mathbb{I}^2$  by dividing it into  $G$  rectangles. The only information retained after the coarse graining is the number of particles in each rectangle. The missing information is the way the particles have been distributed,  $n_1, n_2, \dots, n_G$ , with  $\sum n_k = N$ , the total number of particles. Using Stirling's approximation for the appropriate combinatorial coefficient, this gives for the entropy

$$\frac{S}{kN} = - \sum_{k=1}^G p_k \log p_k \quad \text{with} \quad p_k \equiv \frac{n_k}{N}. \quad (28)$$

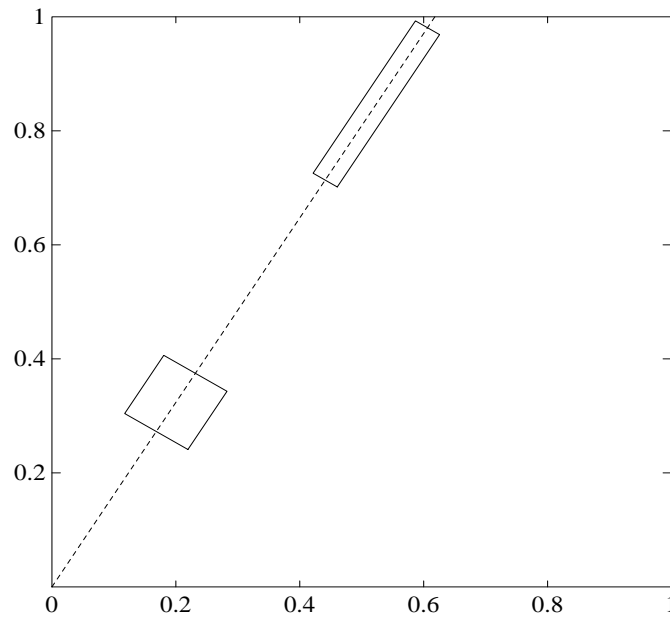


FIG. 5: Action of the cat map on a rectangle.

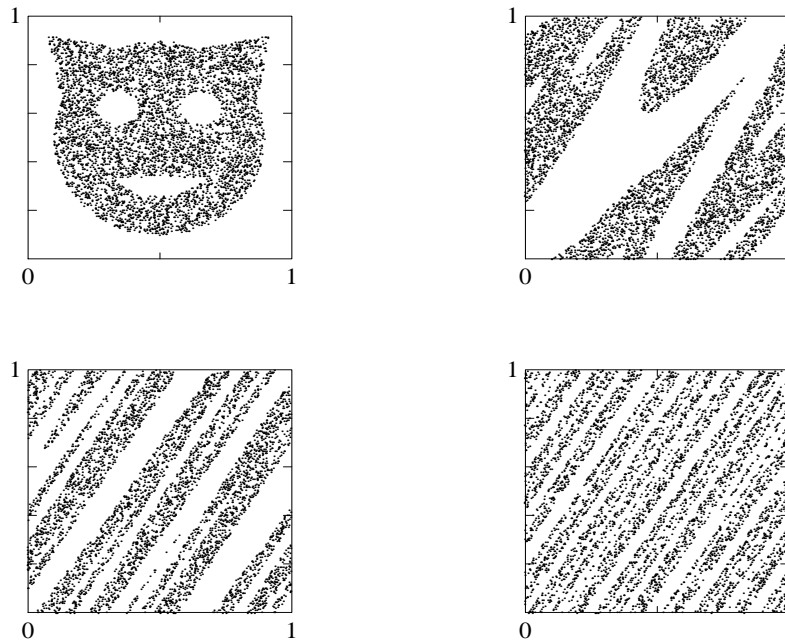


FIG. 6: Action of the cat map on a cat.

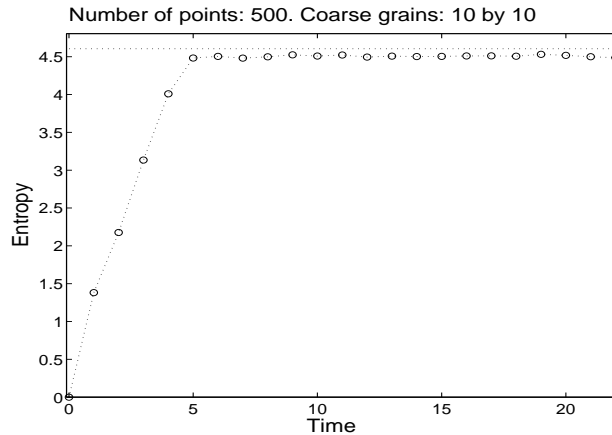


FIG. 7: Entropy as a function of time for the cat map.

(Note: this is *not* the Kolmogorov entropy.)

Generally speaking, entropy increases when energy sequestered in a single macroscopic degree of freedom (or a small number) gets spread among many. For our coarse graining, “macroscopic” means that you can distinguish the various rectangles, but can’t do better than that. So if all the particles (and their energy) go from being in a single rectangle to being spread among many, the entropy has increased. This is shown in Fig. 7.

### 1. Criteria for equilibration

A characteristic of a dynamical system that indicates a fairly strong tendency to equilibrate is *ergodicity*. In the mathematical literature this property comes in many flavors, strong, weak, etc. But I won’t aim for this level of precision.

Generally, ergodicity means phase space averages are equal to time averages. For  $f$  a function on  $\Omega$ , the phase space, the phase space average is

$$\mathcal{P}(f) = \int_{\Omega} f(\omega) d\mu(\omega) \quad (29)$$

The time average needs another argument, namely the initial point on which the time transformation ( $\phi_t$ ) began to act. Thus

$$\mathcal{T}(f, \omega) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T f(\phi_t(\omega)) dt. \quad (30)$$

To say that  $\phi_t$  is ergodic means that  $\mathcal{P}(f) = \mathcal{T}(f, \omega)$ , independent of the initial point,  $\omega$ . This turns out to be equivalent to the assertion that any function invariant under  $\phi_t$  is a constant on  $\Omega$ . See [1] for the proof. (For classical mechanics this constant would be the energy.)

The cat map is even better at coming to equilibrium than merely being ergodic. It is what is called “mixing.” A transformation is mixing if for two (measurable) subsets  $A$  and  $B$  of  $\Omega$ ,

$$\lim_{t \rightarrow \infty} \mu(\phi_t(A) \cap B) = \mu(A)\mu(B), \quad (31)$$

where, as usual, we take  $\mu(\Omega) = 1$ . This represents a kind of statistical independence. Compare it to the formula for conditional probabilities,

$$\Pr(C|D) = \frac{\Pr(C \cap D)}{\Pr(D)}, \quad (32)$$

where  $C$  and  $D$  are “events,” i.e., subsets of  $\Omega$ . When they are independent,  $\Pr(C \cap D) = \Pr(C) \Pr(D)$ , so that in that case,  $\Pr(C|D) = \Pr(C)$ . Belonging to  $D$  gives no information.

To be more explicit, identify probability ( $\Pr$ ) and measure ( $\mu$ ). Then if  $\phi$  is mixing,  $\Pr(\phi_t(\omega) \in A | \omega \in B) = \mu(\phi_t(B) \cap A) / \mu(B) \rightarrow \mu(A)$ .

**Exercise:** Show that mixing implies ergodicity. Hint: use the fact that there is only one constant invariant under  $\phi_t$  for an ergodic system. (You might want to prove that too.) See [1].

### G. The cat map is mixing

This is another digression, but it happens that it does not take much machinery to really prove the mixing property for the cat map, leading to one less piece of wisdom that you are expected to take on faith. I also like this proof because you see that the Fibonacci numbers appear in the cat map story. (Hint:  $\lambda_+$  is exactly 1 more than the Golden Ratio.)

We use the notation,  $M, \xi \equiv \begin{pmatrix} x \\ y \end{pmatrix}$ ,  $\phi$ , etc. of Sec. I F. For measurable sets  $A, B \subset \Omega$ , and  $\phi_t = \phi^t$  for integer  $t$ , we want to prove Eq. (31). Define the characteristic function

$$\chi_A(\xi) \equiv \begin{cases} 1 & \text{if } \xi \in A \\ 0 & \text{if } \xi \notin A \end{cases} \quad (33)$$

Then what we need to show is

$$\lim_{t \rightarrow \infty} \int \chi_{\phi_t(A)}(\xi) \chi_B(\xi) = \int \chi_A(\xi) \int \chi_B(\xi) \quad (34)$$

(suppressing the  $d\mu$ ). The proof uses Fourier analysis and we define the functions

$$e_{pq}(x, y) \equiv e^{2\pi i(px+qy)}, \quad p, q \text{ integers.} \quad (35)$$

We also use the notation  $\nu \equiv \begin{pmatrix} p \\ q \end{pmatrix}$ , so that these functions can be written  $e_\nu(\xi)$ . Then for a measurable function  $f$  on the unit square one has

$$f(x, y) = \sum_{p, q} \hat{f}_{pq} e_{pq}(x, y) = \sum_{\nu} \hat{f}_\nu e_\nu(\xi) \quad (36)$$

with  $\hat{f}_\nu = \int_{\mathbb{T}^2} d^2\xi f(\xi) e_\nu^*(\xi)$ .

The essence of the proof lies in the observation

$$f(\phi(\xi)) = \sum_{\nu} \hat{f}_\nu e_\nu(\phi(\xi)) = \sum_{\nu} \hat{f}_\nu \exp(2\pi i[p(x+y) + q(x+2y)]) \quad (37)$$

$$= \sum_{\nu} \hat{f}_\nu \exp(2\pi i[(p+q)x + (p+2q)y]) \quad (38)$$

$$= \sum_{\nu} \hat{f}_\nu e_{\phi(\nu)}(\xi) = \sum_{\nu} \hat{f}_{\phi^{-1}(\nu)} e_\nu(\xi). \quad (39)$$



Note that in acting on  $\nu$ ,  $\phi$  does *not* invoke the mod 1 operation and is only multiplication by the matrix  $M$ , or  $M^{-1}$  for  $\phi^{-1}$ . The inverse of  $M$  is

$$M^{-1} = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}, \tag{40}$$

which has the same eigenvalues as  $M$ . Let  $u_{\pm}$  be the eigenvectors of  $M^{-1}$  with eigenvalues  $\lambda_{\pm}$  (same  $\lambda$ 's as before, i.e.,  $\lambda_+ > 1$ ). Then we can write

$$\begin{pmatrix} p \\ q \end{pmatrix} = au_+ + bu_-, \tag{41}$$

so that the action of  $\phi^{-1}$  on  $\nu$  is simply multiplication of the appropriate coefficient by the appropriate eigenvalue.

What makes the proof work is that for all  $\nu$ , except  $(0, 0)$ ,  $\phi^{-t}(\nu)$  eventually becomes large. This is because of the (irrational)  $\sqrt{5}$  that appears in  $\lambda$  and  $u$ . Because  $\nu$  consists of integers, *both a and b must be nonzero*.

The functions to which we apply these results are the characteristic functions defined earlier,

$$\chi_A(\xi) = \sum_{\nu} a_{\nu} e_{\nu}(\xi) \quad \text{and} \quad \chi_B(\xi) = \sum_{\nu} b_{\nu} e_{\nu}(\xi). \tag{42}$$

Note that  $a_0 = \int \chi_A = \mu(A)$ , and similarly for  $B$ . Since

$$\xi \in \phi(A) \Leftrightarrow \phi^{-1}(\xi) \in A, \tag{43}$$

$$\chi_{\phi^{-1}(A)}(\xi) = \chi_A(\phi^{-1}(\xi)) = \sum_{\nu} a_{\nu} e_{\nu}(\phi^{-1}(\xi)) = \sum_{\nu} a_{\phi(\nu)} e_{\nu}(\xi), \tag{44}$$

where the last step follows as in the demonstration going the other way ( $\phi$  rather than  $\phi^{-1}$ ). What we need to show is that

$$\Delta_t \equiv \int \chi_{\phi_t(A)}(\xi) \chi_B(\xi) - \int \chi_A(\xi) \int \chi_B(\xi) \rightarrow 0 \tag{45}$$

for  $t \rightarrow \infty$ . This can be written

$$\Delta_t = \int d^2\xi \sum_{\nu} a_{\phi_t(\nu)} e_{\nu}(\xi) \sum_{\mu} b_{\mu} e_{\mu}(\xi) - a_0 b_0. \tag{46}$$

In the usual spirit of convergence proofs, we imagine that some small  $\epsilon$  has been given and we must show that  $|\Delta_t|$  is less than  $\epsilon$  for large enough  $t$ . To do this, truncate the sum over  $\mu$  so that  $\|\mu\|$  is less than some number  $R$  with error (in the sum) less than  $\epsilon$ . By the convergence of the Fourier expansion for the  $\chi$ 's such an  $R$  exists. ( $\|\nu\| = \sqrt{p^2 + q^2}$ .)

By the orthogonality of the Fourier expansion functions, the only non-vanishing terms in the expansion Eq. (46) (now truncated to  $\|\mu\| < R$ ) are those for which  $\mu + \phi_t(\nu) = 0$ . But except for  $(0, 0)$ , for sufficiently large  $t$  all  $\phi_t(\nu)$ 's escape the circle and the sum vanishes. As for the  $(0, 0)$  term, it is cancelled by  $a_0 b_0$ .

Note that the powers of  $M$  give the Fibonacci numbers, as do the powers of  $M - 1$ . The larger eigenvalue of  $M - 1$  is the famous Golden ratio,  $(1 + \sqrt{5})/2$ .

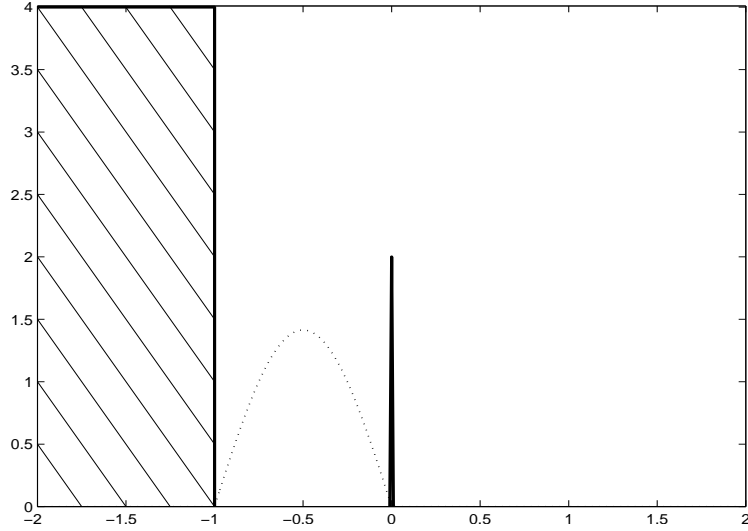


FIG. 8: Arena for quantum decay.

## II. QUANTUM IRREVERSIBILITY

Let a particle be subject to the Hamiltonian

$$H = \frac{p^2}{2} + \lambda\delta(x) + V_{\text{wall}}, \quad (47)$$

where  $V_{\text{wall}}(x)$  is  $\infty$  for  $x < -1$  and zero otherwise. (See Fig. 8.) This has no bound states. Initially we take the wave function to be  $\psi(x, t) = -\Theta(-x)\sqrt{2}\sin(x\pi)$ , and see what becomes of it.

To expand  $\psi$  in eigenstates we solve the time independent problem, whose solution takes the form

$$\psi(x) = \begin{cases} A \sin k(x+1) & \text{if } x < 0 \\ B e^{ikx} & \text{if } x > 0 \end{cases}, \quad (48)$$

where  $E = k^2/2$ . At  $x = 0$  we demand continuity of  $\psi$  and

$$\psi'(0^+) - \psi'(0^-) = 2\lambda\psi(0). \quad (49)$$

This implies

$$k \left( i - \frac{1}{\tan k} \right) = 2\lambda. \quad (50)$$

It is clear that this forces  $k$  to be complex. For large  $\lambda$  there will be metastable states with  $k \approx n\pi$ . We focus on  $n = 1$  and set  $k = \pi + z$  with  $z$  small. The equation satisfied by  $z$  is

$$\frac{2\lambda}{\pi + z} = i - \frac{1}{\tan z}. \quad (51)$$

It's a bit tricky to get the imaginary part of  $z$  here since it turns out to be  $O(1/\lambda^2)$ . In any case, after a bit of algebra one obtains that  $\text{Im } E = -\pi^2/\lambda^2$ . There is thus a decay rate  $\Gamma = 2\pi^2/\lambda^2$ . This system shows irreversibility [10].

Note that  $\text{Im } k < 0$ . This has an interesting implication for the wave function:

$$|\psi(x)| \sim |e^{i(\text{Re } k + i \text{Im } k)x}| = e^{-x \text{Im } k} \rightarrow +\infty \quad \text{for } x \rightarrow \infty. \quad (52)$$

What's going on? This behavior of  $\psi$  emphasizes the inappropriateness of looking at a steady-state solution for the decay problem. We're pretending that this system has been decaying forever. Therefore no matter how small  $\Gamma$  is there was a time in the distant past when there was a lot more probability in the well. The large  $x$  divergence of the wave function thus reflects amplitude that decayed long ago, when there was more amplitude in the interval  $[-1, 0]$ .

Note too where the arrow comes from in this problem. It arises from the outgoing wave boundary condition. We have said, "This is the initial condition: *all* the amplitude is found in the well at  $t = 0$  and we will compute what will happen for  $t > 0$ ."

**Remark:** Recognition of the fact that you cannot evolve arbitrarily far into the past is a feature common to unstable particles. But in that case why is that they have mass and spin as good quantum numbers? For stable particles we have Wigner's analysis of the irreducible representations of the Poincaré group. The labels of these irreducible representation are spin and mass, justifying the use of those quantities for the characterizing of particles. For unstable particles all you have is a *semigroup*. It turns out that you can analyze the irreducible representations of this structure, called the Poincaré semigroup, as well, and can arrive at the same quantum numbers (plus some other representations that you don't get in the stable case). See [11]. This approach has recently been used as a way to more reliably assign masses and widths to unstable particles [12]. It also generates some interesting mathematics connected with finding the states on which this representation acts [13]. See also [14].

To gain perspective on this system's arrow of time, imagine that instead of allowing the particle to escape to infinity we put a hard wall at some distant point  $b$  ( $b \gg 1$ ). We now have a particle in a box, slightly elaborated by having a bump at zero. So we know there is quantum recurrence. Let us estimate the recurrence time.

One can obtain the eigenfunctions of the Hamiltonian (now true, normalizable Hilbert space vectors) by taking the wave function in the interval  $[0, b]$  of the form  $\sin(kx + \phi)$ , but I don't want to go through that exercise. Let the eigenstates of the system, whatever they are, be called  $u_n$ , with energies  $E_n$ . These energies are roughly  $E_n \approx \frac{1}{2} \left(\frac{\pi}{b}\right)^2 n^2$ . So the density of states (" $dn/dE$ ") is essentially  $b^2/n$  or  $b/\sqrt{E}$ . Now start with the same initial wave function as before, the metastable state that looks like the ground state of a particle in a box at  $[-1, 0]$ . We express this wave function in terms of the  $\{u_n\}$ ,

$$\psi(x, 0) = \sum_n c_n u_n(x), \quad (53)$$

for some constants,  $c_n$ . For the time- $t$  state we multiply each term by  $\exp(-iE_n t)$ . For recurrence (as in Sec. IE2) examine the difference

$$\Delta \equiv \|\psi(x, t) - \psi(x, 0)\|^2 = \sum_n |c_n|^2 |1 - \exp(-iE_n t)|^2 = 4 \sum_n |c_n|^2 \sin^2 \left( \frac{E_n t}{2} \right). \quad (54)$$

To make this small we need to look at the range of  $n$  for which  $c_n$  is not small and estimate the probability that all terms  $\sin E_n t/2$  are close zero. We will assume that  $t$  is so large that the numbers  $E_n t/2\pi$  modulo 1 can be considered uniformly distributed on the unit interval. (This is also an assumption about the spectrum, which in a rigorous proof would have to be justified.)

Suppose we want  $\Delta$  smaller than some  $\epsilon$ . We'll use half of  $\epsilon$  to get rid of contributions outside the range of important  $n$ 's. This range can be approximated as the width of this state considered as a resonance, namely the  $\Gamma$  calculated above (this is not a statement about  $\sin \pi x$ , which does not depend on  $\lambda$ , but about the  $\{u_n\}$ , which do). The number of states in the important range of  $n$ 's is thus  $N \equiv \Gamma \cdot [\text{density of states}] = \Gamma dn/dE$ . For simplicity we further assume that the  $c_n$ 's do not vary much in this range, so that each  $|c_n|^2$  is of order  $1/N$ . Therefore to get the sum in Eq. (54) to be less than  $\epsilon/2$  we want each term  $\sin^2 \left( \frac{E_n t}{2} \right)$  be less than  $\epsilon/8$ . This in turn requires that every one of the numbers  $(E_n t/2 \bmod 1)$  lie in an interval of length  $2\sqrt{\epsilon/8}$ . The probability of this occurring is  $\left( \sqrt{\epsilon/2} \right)^{\Gamma dn/dE}$ . Now dropping all order unity constants (which I should have done long ago) and using the estimate above for  $dn/dE$ , this gives  $e^{\Gamma b/\sqrt{E}}$ , or a recurrence time that grows like  $1/\epsilon^b$ . **Exercise:** Show how with  $b < \infty$  the initial packet exits from the confining region. After a short time interval this should look like exponential decay.

A second perspective on quantum decay shows other ways in which irreversibility enters.

Perhaps the simplest model for quantum decay uses the following Hamiltonian and state function

$$H = \begin{pmatrix} \nu & C^\dagger \\ C & \Omega \end{pmatrix}, \quad \psi = \begin{pmatrix} x \\ Y \end{pmatrix}. \quad (55)$$

$\Omega = \text{diag}(\omega_1, \dots, \omega_N)$  (an  $N$ -by- $N$  diagonal matrix with  $\omega_k$  on the diagonal),  $Y$  and  $C$  are  $N$ -component column vectors, and  $\nu$  and  $x$  numbers (so  $H$  is  $(N+1)$ -by- $(N+1)$ ). This is a model for a single level in contact with  $N$  levels and could arise in many contexts (e.g., the Jaynes-Cummings model). Our initial conditions will be that the single level is excited, i.e.,  $x(0) = 1$ ,  $Y(0) = 0$ .

The time-dependent Schrödinger equation (with  $\hbar = 1$ ) is

$$\dot{x} = -i\nu x - iC^\dagger Y, \quad \dot{Y} = -iCx - i\Omega Y. \quad (56)$$

With the stipulated initial conditions, the second equation can be integrated to yield

$$Y(t) = -ie^{-i\Omega t} \int_0^t e^{i\Omega s} Cx(s) ds. \quad (57)$$

This is substituted into the first equation to give a renewal equation for  $x$

$$\dot{x}(t) = -i\nu x(t) - C^\dagger e^{-i\Omega t} \int_0^t e^{i\Omega s} Cx(s) ds \quad (58)$$

For convenience define

$$z(t) \equiv e^{i\nu t} x(t) \quad (59)$$

which produces a quantity that would be entirely stationary, but for the interaction. The equation satisfied by  $z$  is

$$\begin{aligned} \dot{z}(t) &= - \int_0^t e^{i\nu t} C^\dagger e^{-i\Omega(t-s)} C e^{-i\nu s} z(s) ds \\ &= - \int_0^t e^{i\nu t} K(t-s) e^{-i\nu s} z(s) ds \quad \text{with} \quad K(u) \equiv C^\dagger e^{-i\Omega u} C \end{aligned} \quad (60)$$

Everything is still exact. The function,  $K(u)$ , appearing in Eq. (60) with argument  $t-s$ , is of central importance. It is the product of three operators. First  $C$  acts on  $x$  (cf. Eq. (56)) and brings amplitude into the subspace of decayed states. The amplitude in each mode then evolves for a time  $u$ , each mode with its own phase (from  $\Omega$ ). Next  $C^\dagger$  sums the amplitude (the matrix product produces a scalar) and sends it back to the undecayed subspace. Here's the point: to the extent that the modes in the decay subspace get out of phase, the sum going back into the undecayed subspace is reduced. In this way,  $K$  acts so as to shrink the norm in the one-dimensional subspace. The reduction will depend on two factors,  $u$  (larger  $u$  allows the modes to get *more* out of phase) and the range of  $\Omega$  over which  $C$  is large. It is helpful to write  $K(u)$  explicitly. The components of  $C$  will be denoted  $c_k$ .  $K$  becomes

$$K(u) = \sum_k |c_k|^2 e^{-i\omega_k u} \quad (61)$$

Since  $N < \infty$  we *know* that this system recurs (by the quantum recurrence theorem). It follows that the continuum approximation for the levels (which we are about to make) is where the mischief of irreversibility lies.

In the continuum limit the quantities  $c_k$  are of order  $1/\sqrt{N}$ , leading to the definition  $\gamma(\omega) \equiv c_k \sqrt{N}$ . To replace the sum over  $k$  by an integral over energy ( $\omega$ , formerly  $\omega_k$ ), we require the density of states  $\rho(\omega) = 1/[N(\omega_{k+1} - \omega_k)]$  (so the  $N$ 's cancel). In the limit,  $K$  becomes

$$K(u) = \int d\omega \rho(\omega) |\gamma(\omega)|^2 e^{-i\omega u}. \quad (62)$$

What happens next depends on the physics. Generally speaking  $K(u)$  drops off with increasing  $|u|$ . Thus if the coupling of this level to the (now) continuum drops off, then by the Riemann-Lebesgue lemma  $K$  will go to zero. If  $\gamma$  has some significant range  $\Delta E$ , then the spread in  $K$  will be  $1/\Delta E$  (assuming  $\rho$  does not cause mischief). All these properties are what make exponential decay ubiquitous. Having  $K$  drop off on some time scale  $\Delta t$  means that the system forgets the details of its past (cf. Eq. (60)), so that on larger time scales the decay is Markovian.

**Remark:** For a single degree of freedom, being Markovian implies exponential decay. Without memory, the next 1 second is like the last 1 second. If the system shrunk by a factor  $e^{-\gamma}$  ( $\gamma > 0$ ) in the last second, what's left will again shrink by  $e^{-\gamma}$ .

With  $K$  in continuum form the equation for  $z$  becomes (with the change of variable  $u = t - s$  in the integral) [15]

$$\dot{z}(t) = - \int d\omega \rho(\omega) |\gamma(\omega)|^2 \int_0^t du e^{i(\nu-\omega)u} z(t-u). \quad (63)$$

Let us now assume that  $z$  does not change much on the time scale for which  $K$  drops to zero. Then we take it out of the integral, allowing the  $u$  integral to be done explicitly. As in Eq. (61), we require  $\text{Im } \omega > 0$  for convergence, so that our equation becomes

$$\dot{z}(t) = -iz(t) \int d\omega \frac{\rho(\omega) |\gamma(\omega)|^2}{\omega - \nu}. \quad (64)$$

To evaluate this we recall the formula

$$\frac{1}{x \pm i\epsilon} = \mathcal{P} \frac{1}{x} \mp i\pi\delta(x) \quad (65)$$

with  $\mathcal{P}$  the principal value and it is understood that Eq. (65) is to be used in an integral over  $x$ . Note that this formula does not require analyticity of the function integrated and can be justified by multiplying numerator and denominator by  $x \mp i\epsilon$ . As indicated,  $\omega$  has a positive imaginary part, so that Eq. (64) becomes

$$\dot{z}(t) = -iz(t) \left[ \mathcal{P} \int d\omega \frac{\rho(\omega) |\gamma(\omega)|^2}{\omega - \nu} - i\pi\rho(\nu) |\gamma(\nu)|^2 \right] \quad (66)$$

$$= -iz(t)\Delta E - z(t)\Gamma/2, \quad (67)$$

with  $\Delta E$  and  $\Gamma$  defined by Eq. (66). We have thus derived exponential decay with a decay rate given by Fermi's Golden Rule. There is also an energy shift.

Where did the arrow of time enter? By now you are sensitive to the introduction of infinities as the source of mischief, and as usual it is the continuum limit that gives an arrow. We took  $\text{Im } \omega > 0$  to be sure  $K$  (as expressed in Eq. (61)) converges, meaning *converges as*  $u \rightarrow +\infty$ . There is an initial state at time zero and the expectation of reasonable behavior for the distant future, with “future” have implicitly a particular sign of  $t$ . Prior to the continuum limit there was both no need to introduce imaginary parts and there was the prospect of ultimate recurrence.

**Remark:** *The quantum Zeno effect.* Return to Eq. (60). Note that as  $t \searrow 0$ ,  $\dot{z}$  goes to zero. So although an undisturbed system ultimately has exponential decay (with  $|\dot{z}/z| = \Gamma/2 > 0$ ), things start out slowly. Suppose one would “check” at a very early time whether the system is still in its original state, than the probability of finding a change would be less than  $O(dt)$ . This is the basis of the quantum Zeno effect (QZE). You repeatedly measure, or project, at sufficiently short intervals  $dt$ , and in this way prevent decay altogether. A watched pot never boils! There is a considerable literature on this subject. An early discoverer of this effect (it seems it was discovered several times independently) was L. Khalfin. One recent theoretical paper is [16]. The effect has also been seen experimentally in [17]. The subject of continuous observation (raised during the lecture) arises because it seems at first sight that if one were to stare at a sample of radium (or maybe hold a Geiger counter to it) it should not decay. The reason that it does decay is that all such measurement devices, Geiger counters, eyeballs, etc., have a time scale. If this time scale is shorter than the time interval needed to produce the QZE, then indeed decay is halted. One can build one's own internal observer by extending the Hamiltonian of Eq. (55) slightly:

$$H = \begin{pmatrix} \nu & \Phi^\dagger & 0 \\ \Phi & \omega & \Theta^\dagger \\ 0 & \Theta & W \end{pmatrix}. \quad (68)$$

$W$ , like  $\Omega$ , is diagonal and represents levels to which the system can decay after it has reached the first second quasi-continuum. So these levels “observe” the transition. If the

coupling  $\Theta$  is very strong then “as soon” as the decay takes place it is noticed. This stops the decay. There is also an energy-level interpretation. See [18].

**Remark:** Again I refer to Eq. (60). This time I note that the function  $\rho|\gamma|^2$  is bounded from below, that is, below some  $E_0$  there are no levels. By general theorems of the Fourier transform, this implies that the asymptotics of  $K$  with  $u$  cannot be exponential. In practice, depending on the threshold form of the function the asymptotic decay behavior must be power law,  $t$  to some negative power. This too was predicted by Khalfin. This has never been observed experimentally, and is expected to be difficult to see [19].

### III. STOCHASTIC DYNAMICS: IRREVERSIBILITY FROM COARSE GRAINING

In this Section I will introduce information-theory concepts that tie in directly to irreversibility. It is interesting that the proofs of the essential relations go through with few assumptions, both for classical mechanics and for quantum mechanics. One feature that will be important is the question of what is a good “coarse graining.” This will also come up later in these lectures, and the present discussion will introduce both the issues and the formalism.

First, classical mechanics. We work on phase space,  $\Omega$ . Let  $\mu$  be the measure on  $\Omega$  and  $\mu(\Omega) = 1$ . Let the dynamics be given by a measure-preserving map  $\phi^{(t)}$  on  $\Omega$ , with  $\phi^{(t)}(\omega)$  the time- $t$  image of an initial point  $\omega \in \Omega$ . The time parameter,  $t$ , may be either continuous or discrete. For simplicity we fix a time  $t$ , and call the associated phase space map,  $\phi$ . The notion of “macroscopic” is provided by a coarse graining on  $\Omega$ . This is a finite collection of sets of strictly positive measure that cover  $\Omega$ :  $\{\Delta_\alpha\}$ ,  $\alpha = 1, \dots, G$ , with  $\cup_\alpha \Delta_\alpha = \Omega$ ,  $\Delta_\alpha \cap \Delta_\beta = \emptyset$  for  $\alpha \neq \beta$ . Let  $\chi_\alpha$  be the characteristic function of  $\Delta_\alpha$  and let  $v_\alpha = \mu(\Delta_\alpha)$ . If  $f$  is a function on  $\Omega$ , its coarse graining is defined to be

$$\hat{f}(\omega) \equiv \sum_\alpha \frac{\chi_\alpha(\omega)}{v_\alpha} f_\alpha, \text{ with } f_\alpha = \int d\mu \chi_\alpha(\omega) f(\omega). \quad (69)$$

Note that  $f_\alpha/v_\alpha$  is the average of  $f$  on  $\Delta_\alpha$ , and that  $\int_\Omega f = \int_\Omega \hat{f}$ . The function  $f$  has been replaced by a step function that is constant on the grains, in such a way that its total integral has been preserved. Eq. (69) can also be written

$$\hat{f}(\omega) = \sum_\alpha \frac{\chi_\alpha(\omega)}{v_\alpha} \int d\mu(\omega') \chi_\alpha(\omega') f(\omega'). \quad (70)$$

Let the system’s distribution in  $\Omega$  be described by a density function  $\rho(\omega)$ . We take the primitive entropy to be

$$S_{\text{prim}} = - \int_\Omega \rho(\omega) \log(\rho(\omega)) d\mu. \quad (71)$$

Note that this differs by a (possibly infinite) constant from entropy defined on a discrete set—see Eq. (21). This is because discretizing phase space would give  $-\sum p \Delta x \log(p \Delta x)$  as the usual information entropy.  $S$  in Eq. (71) differs from this by  $\log \Delta x$  and as a result may be positive or negative. By virtue of the fact that  $\phi^{(t)}$  is measure preserving and invertible,  $S_{\text{prim}}$  is constant in time.

The entropy that I will show to be non-decreasing is the *coarse-grained* entropy, and is defined as

$$S(\rho) \equiv S_{\text{prim}}(\widehat{\rho}) = - \int_{\Omega} \widehat{\rho} \log \widehat{\rho} d\mu. \quad (72)$$

with  $\widehat{\rho}$  formed from  $\rho$  as in Eq. (69). By the definitions,

$$S(\rho) = - \int d\mu(\omega) \sum_{\alpha} \frac{\rho_{\alpha}}{v_{\alpha}} \chi_{\alpha}(\omega) \log \left( \sum_{\beta} \frac{\rho_{\beta}}{v_{\beta}} \chi_{\beta}(\omega) \right) \quad (73)$$

$$= - \sum_{\alpha} \rho_{\alpha} \log \left( \frac{\rho_{\alpha}}{v_{\alpha}} \right) \quad (74)$$

$$= S(\rho_{\alpha}|v_{\alpha}), \quad (75)$$

where the second line follows since there is no contribution unless  $\alpha = \beta$ ; in that case all one gets is integration over particular grains. On each grain  $\widehat{\rho}$  is constant and the integral over  $\chi_{\alpha}$  yields  $v_{\alpha}$ , cancelling the one inside the sum.

The function  $S(p|q)$  is the *relative entropy* defined by

$$S(p|q) \equiv - \sum_x p(x) \log \left( \frac{p(x)}{q(x)} \right), \quad (76)$$

where  $p$  and  $q$  are probability distributions such that  $q(x)$  vanishes only if  $p(x)$  does.  $S(p|q)$  is also known as the Kullback-Leibler distance. Note that  $\sum \rho_{\alpha} = \int \rho = 1$ , and that all  $v_{\alpha} > 0$ . We will use two properties of relative entropy that I will not prove:

$$S(p|q) \leq 0 \quad (77)$$

$$S(p|q) \leq S(Rp|Rq) \quad (78)$$

where  $R$  in Eq. (78) is a *stochastic* matrix. A stochastic matrix has the following two essential properties:  $R_{xy} \geq 0$  and  $\sum_x R_{xy} = 1$ , for all  $y$  (and  $x$  and  $y$  are the matrix indices). The significance of such a matrix is that it represents the probability of a transition from  $y$  to  $x$  in unit time for a Markov process on a set  $X$ , with  $x, y \in X$ . (Some authors define  $R$  the other way round.) These properties are proved in [20, 21] and also in [22].

What I will now show is that  $S(\widehat{\phi(\widehat{\rho})}) \geq S(\widehat{\rho})$ . In other words, if you take the coarse-grained distribution function forward in time and again coarse grain it, entropy increases.

We already know  $S(\widehat{\rho})$ , it is just  $S(\rho_{\alpha}|v_{\alpha})$  (there is a slight abuse of notation in using  $\alpha$  in both arguments of  $S$ , but let's not get fussy). We next examine the coarse graining of the time image of  $\widehat{\rho}$ . To avoid enormous "hat" symbols, I write  $\widehat{a} = [a]_{\text{c.g.}}$ .

$$[\widehat{\phi(\widehat{\rho})}]_{\text{c.g.}}(\omega) = \sum_{\alpha} [\chi_{\alpha}(\phi(\omega))]_{\text{c.g.}} \frac{\rho_{\alpha}}{v_{\alpha}} \quad (79)$$

$$= \sum_{\alpha} \sum_{\beta} \frac{\rho_{\alpha}}{v_{\alpha} v_{\beta}} \chi_{\beta}(\omega) \int d\mu(\omega') \chi_{\beta}(\omega') \chi_{\alpha}(\phi(\omega')) \quad (80)$$

$$= \sum_{\beta} \frac{\chi_{\beta}(\omega)}{v_{\beta}} \frac{\mu(\Delta_{\beta} \cap \phi^{-1}(\Delta_{\alpha}))}{v_{\alpha}} \rho_{\alpha} \quad (81)$$

$$\equiv \sum_{\beta} \frac{\chi_{\beta}(\omega)}{v_{\beta}} \rho'_{\beta}, \quad (82)$$



TABLE I: Classical-quantum correspondence for the entropy increase proposition.

classical	quantum	classical	quantum
$\Omega$	$\mathcal{H}$ (Hilbert space)	$\rho$	$\rho$ (density matrix)
$\Delta_\alpha$	$\mathcal{H}_\alpha$ (subspace)	$\mu$	Trace
$\chi_\alpha$	$P_\alpha$ (projector)	$v_\alpha$	dimension of $\mathcal{H}_\alpha$
$\phi^{(t)}$	$U_t$		

where Eq. (80) uses Eq. (70). Eq. (82) defines the new coefficients  $\rho'$  which we can write as

$$\rho'_\beta = \sum_\alpha R_{\beta\alpha} \rho_\alpha, \quad (83)$$

with

$$R_{\beta\alpha} \equiv \frac{\mu(\Delta_\beta \cap \phi^{-1}(\Delta_\alpha))}{v_\alpha}. \quad (84)$$

The proof can now be easily completed by noting first that  $R$  is stochastic and second that it has eigenfunction  $v_\alpha$  with eigenvalue unity. The proof of stochasticity follows by adding all  $\beta$  and, since  $\cup_\beta \Delta_\beta = \Omega$ , the result is one. The eigenvector property is similarly immediate. It follows that the entropy of  $[\phi(\hat{\rho})]_{\text{c.g.}}$  is  $S(\rho'|v_\alpha) = S(R\rho|Rv_\alpha) \geq S(\rho|v)$ .

It is remarkable that this proof goes through for any partition. In particular, you might have thought that the grains through which a system passes as it moves toward equilibrium must get larger and larger. This would be in keeping with the idea that the system is *really* in some single grain and its entropy is the logarithm of that grain's volume. That may happen but it need not, since entropy increase occurs through uncertainty over which grain it's actually in. Put otherwise, the increase occurs by virtue of the density's *spreading* to many grains, irrespective of their individual sizes.

Where then does the physics enter? It is in the assertion that it is physically meaningful to coarse grain the image of the coarse-grained  $\rho$ . The implicit assumption is that *within each grain* the phase space points have spread uniformly. Thus  $t$  cannot be arbitrarily small and must exceed a microscopic relaxation time associated with the coarse grains.

**Exercise:** This can also be phrased in terms of the probabilities that the  $\phi$ -image of a subset of  $\Delta_\alpha$  goes to the grains in the image of  $\Delta_\alpha$  in the same proportion that  $\Delta_\alpha$  itself does.

When I first worked out this little calculation I thought that for the quantum case things would be more complicated, with all sorts of phases entering and giving the desired result only if they miraculously cancelled. This turns out not to be the case and the quantum version is as straightforward as its classical counterpart.

I will not go through the details. One only needs a dictionary, as given in Table I. Thus,

$$\hat{\rho} = \sum \frac{P_\alpha}{v_\alpha} \rho_\alpha, \quad \text{with } \rho_\alpha = \text{Tr} P_\alpha \rho \text{ and } v_\alpha = \text{Tr} P_\alpha. \quad (85)$$

Entropy is again  $S(\rho_\alpha|v_\alpha)$ . (The  $v$ 's no longer sum to unity, but this makes no difference.) Time evolution is given by a unitary operator,  $U_t$ , acting in the usual way:  $\rho(t) = U_t \rho(0) U_t^\dagger$ . Carrying through the same steps as for the classical case, coarse graining, evolving in time

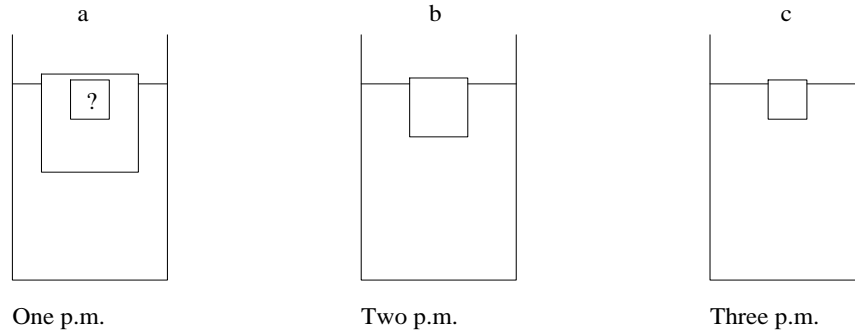


FIG. 9: Three images of a glass of water with an ice cube in it.

and coarse graining again, leads to the same equations, but with the matrix “ $R$ ” now given by

$$R(\beta, \alpha) = \text{Tr} \left[ P_\beta U_t P_\alpha U_t^\dagger \right] / v_\alpha. \quad (86)$$

Stochasticity of  $R$  is readily established and entropy non-decrease follows as above.

Exercise: Check that  $R$  is real.

#### IV. INITIAL CONDITIONS VS. FINAL CONDITIONS: A RESTATEMENT OF THE SECOND LAW

Given the present-time macrostate of a system, the second law of thermodynamics can be phrased as a distinction between the way we compute the later state of the system, and the way we estimate its earlier state. These calculations are called *prediction* and *retrodiction*.

Consider the situation pictured in Fig. 9. I show three images of a glass of water with an ice cube in it. Suppose you are told that at 2 p.m. the water is at temperature  $T_2$  and that the ice cube volume is  $V_2$ , as shown in the center image. The system is isolated between 2 p.m. and 3 p.m. (the walls of the larger container are not shown) and you are asked to predict the water temperature and ice cube size at 3 p.m..

Your prediction should look like Fig. 9c: a smaller ice cube and colder water,  $V_3 < V_2$ , and  $T_3 < T_2$ . In principle, how did you do your calculation? The middle picture (9b) is a *macroscopic* description and has an associated a region of phase space, call it  $\Gamma_2$ . Every microstate  $\gamma \in \Gamma_2$  looks like Fig. 9b. To make your 3 p.m. prediction, you evolve each  $\gamma$  forward by one hour. The vast majority will look like Fig. 9c, and this is your predicted macrostate. Explicitly stated: you average over the microscopic states of  $\Gamma_2$ , and give them all the same weight. This reflects a fundamental principle of statistical mechanics. For example, for an isolated system (as this one is) the partition function sums equally over all states.

This procedure can be described in phase space. In Fig. 10, the oval in the middle represents the phase space region  $\Gamma_2$ . From  $\Gamma_2$  all microstates begin at 2 p.m.. Let the operator mapping phase space points forward one hour be  $\varphi$ . The statement about equal *a priori* probabilities at 2 p.m. is the statement that our 2 p.m. phase space measure is the characteristic function of  $\Gamma_2$ . By 3 p.m. the system has evolved to  $\varphi(\Gamma_2)$ . (See Fig. 10.) This is no longer a macroscopic state. The image of a coarse grain need *not* be a coarse grain. The typical situation is that  $\varphi(\Gamma_2)$  falls into one or several coarse grains at the

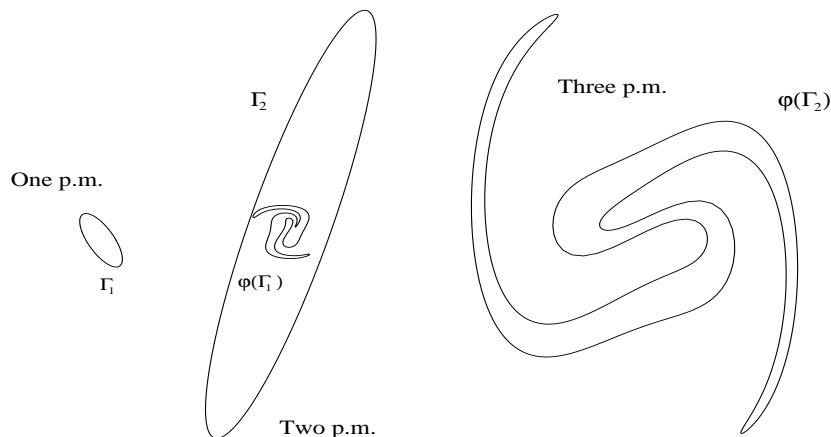


FIG. 10: Phase space. Schematic illustration of  $\Gamma_2$ ,  $\varphi(\Gamma_2)$  (the dynamical image of  $\Gamma_2$  after 1 hour),  $\Gamma_1$ , and  $\varphi(\Gamma_1)$ .

later time. Sometimes the image grains will be far apart, for example for the phase space description of a person throwing dice. The objective of a dice throw is to make what looks like a single macroscopic initial state go into several distinct macroscopic final states. Yet another feature of  $\varphi(\Gamma_2)$  is that small portions of it do not look like the right-hand figure at all. This will be clear later.

The macroscopic state associated with the right-hand picture in Fig. 9 is some subset of phase space, call it  $\Gamma_3$ . (This is not pictured in Fig. 10.) The conclusion to be drawn from the last paragraph is that, to a good approximation,  $\varphi(\Gamma_2) \subset \Gamma_3$ . What a “good approximation” means is that  $\mu(\varphi(\Gamma_2) - \Gamma_3) \ll \mu(\Gamma_2)$ , where  $\mu$  is the phase space measure. So  $\Gamma_2$  has found its way into  $\Gamma_3$ , but how much of  $\Gamma_3$  does it occupy? The answer is, very little. The phase space volume  $\Gamma_2$  has been stretched, whirled and tendrilled so as to occupy  $\Gamma_3$ , all the while maintaining its original phase space volume  $\mu(\Gamma_2)$ . This connects our discussion to the second law of thermodynamics. The entropy at 2 p.m. is the logarithm of the volume of  $\Gamma_2$ . The entropy at 3 p.m. is the logarithm of  $\mu(\Gamma_3)$ . Therefore for a glass of normal size the ratio of the phase space volumes will be on the order of  $\exp(N_A)$ , with  $N_A$  Avogadro’s number. So it is the fact that a coarse grain at one time spreads into many coarse grains, or larger coarse grains, at later times that corresponds to the second law of thermodynamics. Implicit is the assumption that the earlier grain was uniformly occupied, otherwise you couldn’t be sure that all the regions at the later time were also occupied.

Now a more subtle question: You have additional information: the system was isolated from 1 p.m. to 3 p.m., not just from 2 p.m.. What was the system state at 1 p.m.? One way to answer would be to apply  $\varphi^{-1}$  to  $\Gamma_2$  and identify the coarse grains needed to contain most of  $\varphi^{-1}(\Gamma_2)$ . Since  $\varphi$  and  $\varphi^{-1}$  are the same, or essentially the same, the 1 p.m. macrostate of the system would be the smaller cube shown on the left of Fig. 9 (along with colder water).

Of course it doesn’t happen that way. Had you started with a small cube at 1, you would have an even *smaller* one at 2. Obviously there is a different rule for retrodicting. What is it?

There is no unique answer. Here is one way. Make a hypothesis, about the 1 p.m. macrostate. This corresponds to some set  $\Gamma_1$  in phase space (illustrated on the left of Fig. 10). Propagate it forward one hour to get  $\varphi(\Gamma_1)$ . If this lies substantially within  $\Gamma_2$  (as it does in Fig. 10), then your hypothesis was acceptable. For example, the larger cube on

the left of Fig. 9, might be a precursor to the middle picture. But other precursors are also possible. For example, you might have had a smaller cube plus a little shaving of ice that melted entirely during the hour. The coarseness of your coarse grains may preclude your being able to discriminate between these cases. You may even have several hypotheses and want to use other information to weight your guesses. For example you may know that all available ice came from a machine whose maximum cube size is smaller than the large one pictured on the left, implying that there had to be some small chips at 1 p.m. [23].

You do know however that the combined phase space volume of all acceptable hypotheses is vastly smaller than  $\mu(\Gamma_2)$ . This is because most of the volume in  $\Gamma_2$ , when propagated back to 1 p.m., gives the single smaller ice cube, the one with the question mark in it in the figure. So if  $\Gamma_1$  is taken to represent the volume of all acceptable 1 p.m. phase space volumes,  $\mu(\Gamma_1) \ll \mu(\Gamma_2)$ .

I summarize: To predict, take—with equal weight—all microstates consistent with the macrostate and propagate them forward in time. Then coarse grain to get your macroscopic prediction. To retrodict, go back to the earlier time and make macroscopic hypotheses. Use each such hypothetical macrostate to predict the present state. If it agrees, it was a possible precursor. In general, the phase space volume of all such precursors is less than or equal to that of the subsequent macrostate. In this way there is entropy non-decrease for an isolated system when moving forward in time. These different rules for prediction and retrodiction are an alternative statement of the second law of thermodynamics.

Here is yet another statement of the past/future distinction. Suppose you know the 2 p.m. macrostate and are asked what you know about its microstates. If all you are concerned with is prediction, then you say, all microstates are equally likely. But if you happen to know that the system had been isolated since 1 p.m., then the vast majority of those microstates are rejected and all that are allowed are those in the set  $\varphi(\Gamma_1)$ , with  $\Gamma_1$  your collection of acceptable hypotheses. In other words, for an *initial* state you treat all microstates equally, but if something is a *final* state the vast majority of its microstates are rejected.

The foregoing observation shows that the assumptions of statistical mechanics are even stronger than was already stated. One way to make a prediction is to give equal *a priori* probability to all initial microstates. But that's not the only way. For our system isolated from 1 p.m. to 3 p.m., the set of 2 p.m. microstates is now known *not* to be  $\Gamma_2$ , but  $\varphi(\Gamma_1)$ , a much smaller set. That means that our 3 p.m. macrostate need not be calculated from  $\varphi(\Gamma_2)$ , but from  $\varphi(\varphi(\Gamma_1))$ . Does anyone think the 3 p.m. result will be any different? All experience says otherwise. It does not matter whether you use all of  $\Gamma_2$  or only the part that results from a state that had been isolated earlier. Presumably,  $\varphi(\Gamma_1)$  fills  $\Gamma_2$  in a pseudo-random way, so that its points are representative of what *all* of  $\Gamma_2$  is doing.

### 1. Past/future reliability: not a restatement of the second law

The feeling that the past has occurred, is known and unchangeable, while the future is open, is a psychological arrow of time. Here I distinguish this arrow from the notion of reliability. The relative accuracy of prediction and retrodiction is separate issue.

In 1962, Gold (quoting Levinger) gave the example of a Russian satellite. Here prediction is more effective than retrodiction. A satellite in orbit allows accurate prediction of its free trajectory. However, it is difficult to deduce where it was launched. (In those Cold War days, such information was often secret.)

Consider a roulette wheel. Saying that the system is difficult to predict means that the

set  $\varphi(\Gamma_2)$  has significant weight in more than one macrostate. To state this precisely, let  $\Delta_k$  be the phase space grains at time  $T_3$  ( $k$  runs over an index set). Uncertainty then means that  $\Pr(\Delta_k|\varphi(\Gamma_2)) = \mu(\Delta_k \cap \phi(\Gamma_2))/\mu(\Gamma_2)$  is well away from zero for several  $k$ . For a roulette wheel, from the initial condition, you cannot tell into which trap the ball will fall. For a symmetric example, imagine a slightly nonconventional roulette wheel. Prior to time  $T_1$  the ball is at rest in one of the traps. At  $T_1$  the wheel is spun vigorously. It leaves the trap it is in, bounces around, and then (at  $T_3$ ) is caught by another trap. Now you look at the time  $T_2$  position, with  $T_1 < T_2 < T_3$ . You don't know where it came from and you don't know where it is going. Call the  $T_1$ -coarse grains  $\delta_k$ . With this notation, your various time  $T_1$  hypotheses would be realized with probability proportional to  $\mu[\Gamma_2 \cap \phi(\delta_k)]$ , and your uncertainty would correspond to this being relatively large for several values of  $k$ .

*Why then when we remember something does it seem more certain than when we predict?* This is because we are not retrodicting from the system itself, but from an auxiliary system, our brain, on which the earlier state of the system caused modification of neural coarse grains, i.e., wrote to memory. The fact that coarse grains in the brain have relatively few precursors is a property you want for any recording system.

## V. ARROWS OF TIME

- The thermodynamic arrow. Entropy increase, heat cannot be converted to work, rules for predicting and retrodicting (with caveats, modifiers, etc.)
- The biological arrow.
  - Digestion. Presumably follows the thermodynamic arrow
  - The past is accomplished, done. The future is changeable. This is a tougher nut, if you want to attribute it to the thermodynamic arrow. I have tried to narrow the issues by considering the following notion: “A computer’s arrow of time” [24]. Imagine a weather-predicting computer. It has long term memory, records of external weather as well as its own calculations. It has short term memory, it’s own current calculations, perhaps non-archived new data. (On a PC this could be the hard drive-RAM distinction.) It may have sensory organs, direct data feeds. It may be able to seek more data, sending up a weather balloon or instructing a human to do so. It may also be able to act, ordering the seeding of clouds, based perhaps on goals that have been programmed into the machine. It certainly deals with the past and future in different ways, but one would nevertheless ascribe all these asymmetric phenomena to the thermodynamic arrow.

*Which leads me to recall the last bastion of the biological arrow:*

- Consciousness. Who knows? This concept is the playground of neurophysiologists, philosophers and the occasional brave physicist [25]. But not (yet?) me. The philosophers use notions that I haven’t been able to figure out, for example, qualia. Perhaps you’ll do better. Here’s Webster: **quale**: *a sense-datum or feeling having its own particular quality without meaning or external reference <a quale... such as, say, a certain shade of purple occurring in the presentation of a certain piece of cloth in a particular light— C.G.Hempel.>*
- Remark: Despite dissipation, the world has gotten more interesting in the last 4 Gyr. It’s because the Earth is an *open* system, the beneficiary of an enormous

negative entropy input. Photons of 6000K arrive from the sun, and assuming a rough energy balance, 20 of them leave at 300K. So the sun's gift to us is not so much energy (which, if it does accumulate, can cause problems), as negentropy. The same is true for computers, some of whose power goes into a fan, whose job it is to remove energy from the overheated chip. I believe the first to appreciate this was Schrödinger [26].

- The CP arrow. Given the observed CP violation, CPT implies T asymmetry. So the laws of physics are presumably fundamentally time asymmetric.

I don't believe this asymmetry is relevant to the thermodynamic arrow. Having different dynamics in one direction and the other should not make a difference (at least the kind of difference we see in the Second Law) provided the laws induce equilibrium in roughly the same way (and are not intrinsically dissipative).

In Sec. VII B, when more tools for assessing the impact of asymmetry are developed, I'll give a cat map-based example that supports this opinion. I also remark that in the early 1970's there were some attempts to relate the thermodynamic arrow to the CP arrow, but none as far as I know met with success.

- The cosmological arrow. The universe is expanding.
- The radiative arrow. Variouslly defined: one uses outgoing wave boundary conditions for electromagnetic radiation; retarded Green's functions are used for ordinary calculations; radiation reaction has a certain sign; more radiation escapes to the cosmos than comes in. Because the theory is time-symmetric these phenomena have reverse-time descriptions also, so the arrow is not precise. Let me single out two candidates.
  - Radiation reaction. It costs you energy to shake a charged particle. I would relate this to the thermodynamic arrow. It also costs you energy to stir your coffee, but only if the coffee began at rest.
  - Escape of radiation to the cosmos. Olber's paradox suggests this is a consequence of the cosmological arrow.

## VI. CORRELATING ARROWS OF TIME

Around 1960 Thomas Gold argued that the thermodynamic arrow is a consequence of the cosmological arrow. Your coffee cools because the galaxies recede. He said that it's because expansion allows the universe at large to serve as a "sink" in the thermodynamic sense. He argued that in this way the global could impose an arrow on the local. In the following argument he assumes that the radiative arrow (radiation leaves us, "never" to return) is a consequence of the cosmological arrow and shows by example how that can impose a local arrow. Here's the argument in Gold's words [27]:

Let us take, for example, a star, and suppose we could put it inside an insulating box. . . . when the star has been in the box for long enough (which in this case will perhaps be rather long), time's arrow will have vanished. . . . now if we were to open for a moment a small window in our box, then what would happen? Time's arrow would again be defined inside the box for some time, until the statistical equilibrium had been reestablished. But what had happened

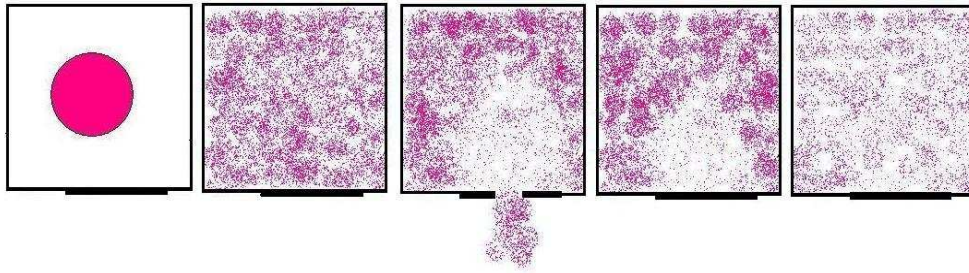


FIG. 11: Illustration for Gold's star-in-a-box parable. Five snapshots.

when we opened the hole? Some radiation had, no doubt, escaped from the box and the amount of radiation that found its way into the box from the outside was incomparably smaller.

The escape of radiation away from the system is, in fact, characteristic of the type of “influence” which is exerted from outside. ... The thermodynamic approach would be to explain that ... free energy can only be generated from the heat sources in the world by means of heat engines working between a source and a sink. There may be a variety of sources, but the sink is always eventually the depth of space, ...

It is this facility of the universe to soak up any amount of radiation that ... enables it to define the arrow of time in any system that is in contact with this sink.

In Fig. 11 I have illustrated five stages in this story. The star, which has some base value of entropy, is initially not much affected by the confinement. As the star ages, it will tend to equilibrate within the box. I won't worry about details, like finding a box that confines neutrinos as well as everything else, and will assume that eventually the box is filled uniformly by very hot gases, including photons. As with all equilibration processes, the entropy will rise in the many-billion-year course of this process. Then the window is opened and photons *escape*, rather than enter. The entropy in the box decreases during the short interval when the window is open, but once closed equilibration sets in again and an entropy maximum is again reached. This entropic history is shown, slightly mismatched horizontally, in Fig. 12, with a better view of the entropy alone in the upper portion of Fig. 13. The time units for the figure take 0 as the start of the story, 100 as the time the window is opened, 105 when it is closed.

The absolutely essential point, on which Gold basis his conclusion, is that during the recovery period—after time 105—the entropy climbs, presumably to a bit below its previous equilibrium value. Note that for  $t > 105$  entropy is *increasing*, whereas for times less than 100 it was constant (except of course for fluctuations, also schematically indicated). In this way, the universe, and in particular the outgoing wave boundary condition, which determined that the time 100 to 105 flux was *outward*, has imposed its arrow on the local region, i.e., within the box. Our coffee cools because we are contact with the universe, *and it is expanding*.

Does this show how we inherit the thermodynamic arrow from the cosmological arrow? Unfortunately, not. Suppose that when you opened the box radiation entered rather than

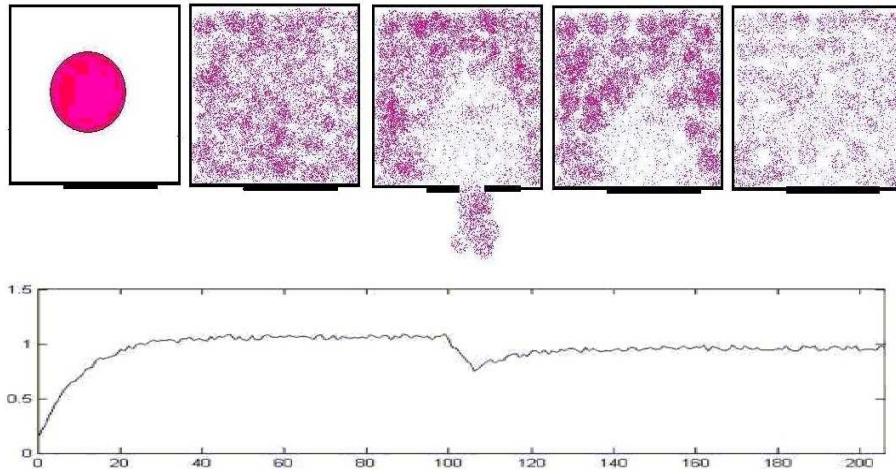


FIG. 12: Star-in-a-box story with schematic view of entropy dependence.

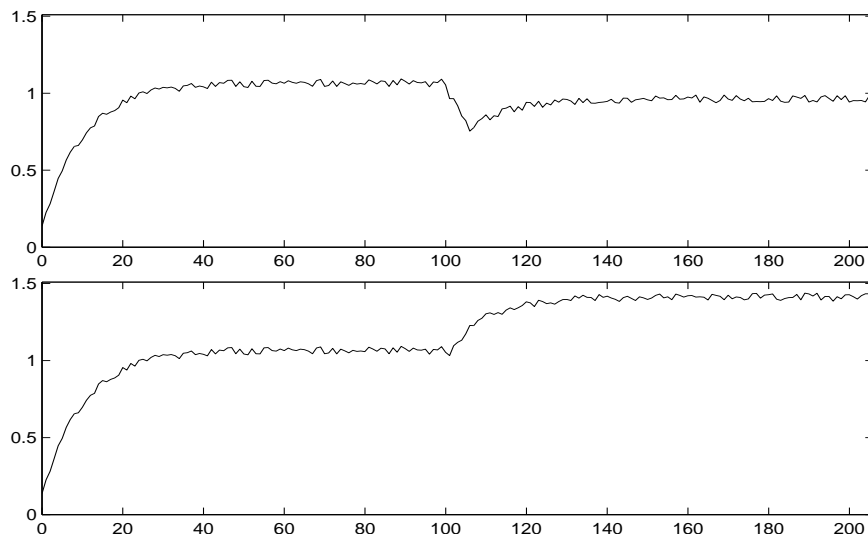


FIG. 13: Possible time dependence of entropy for the star-in-a-box. Upper graph: as expected from the star-in-a-box parable. Initially entropy increases as the system approaches equilibrium. At time 100 (in the figure units) the window is opened and the entropy drops. After closing the window (at time 105), equilibrium is restored. Only the entropy *inside* the box is shown. Lower graph: Same as before, except that radiation is assumed to *enter* during the time interval [100, 105].

left. Just as would occur for *departing* radiation, the system will be thrown out of equilibrium. The final entropy might be larger (see Fig. 13), but the essential feature, the fact that the outside influence reestablished an arrow of time in the box, would still hold.

This counterexample does not contradict the importance of outside factors in imposing a local thermodynamic arrow, but it does show that the arrow is not necessarily a consequence of the *outgoing* property.

It is instructive to ask, where did the arrow that we “derived” actually come from? Why did we naturally accept that the behavior should resemble that in Fig. 13? Why should the



return to equilibrium *follow* the opening of the window? The answer is that the arrow of time in Gold’s argument is the arrow of time of the *narrator*. In telling the story it is natural for us to assume that the response of the system follows, in our time sense, the opening of the window. This is the same assumption that is involved in the use of initial conditions in the posing of macroscopic, dissipative problems.

This is an example of how, in discussing the arrow of time, ordinary language presents us with traps. In the foregoing argument the “natural” use of *initial* conditions is what gave the arrow. As discussed earlier, this use of macroscopic initial conditions is another way of formulating the thermodynamic arrow of time. So in the end the argument is circular.

What is the alternative? If initial conditions prejudice the answer, don’t use them. If the claim is that the thermodynamic arrow is a consequence of the expansion of the universe, then one way to handle the problem would be to give partial boundary conditions at two different times. At one time the universe is in a contracted state, at the other, it is much larger. Then you would try to give the same kind of information at those two times, so that it would only be the expansion that could give rise to the arrow. Even better, since it might not be clear what “same kind of information” means, you might consider an oscillating universe and then give the same kind of information for times a bit after the big bang and a bit (the same bit) before the big crunch. For example, you might propose local equilibrium 5 minutes into history and 5 minutes before the end of history.

## VII. THE UNIVERSE WITH TWO-TIME BOUNDARY CONDITIONS

For the sake of the logic, and not necessarily reflecting our actual cosmology, we assume a roughly time-symmetric universe, with expansion and contraction. In our expanding phase there was an important transition: In early times short-range forces dominated and the universe was homogenous (evident in the cosmic background radiation distribution). This was the highest entropy state, the most likely macroscopic state of the universe, given the forces governing the motion of matter at the time.

On the other hand, with rapid expansion, the matter separated into clumps, with distant clumps no longer interacting by short range forces, but rather by gravity. Gravity does not favor homogeneity. With time the local configuration separated into stars, galaxies, clusters of galaxies, etc. So what *was* the highest entropy state—homogeneity—becomes an extremely unlikely state. You’ll notice I avoid assigning an entropy or using the word equilibrium, since with long range forces those concepts become problematic. Nevertheless, once gravity becomes the dominant force, the system finds itself in what would be an unlikely state had gravity been dominant before. If we wanted to model this in a non-gravitational context it would be reasonable to consider two-time boundary conditions with low entropy macroscopic states at each end.

### 1. *How the universe imposes its arrow.*

*The foregoing argument is the decisive one, even if you don’t want to talk about two time boundary conditions.* Boltzmann, in trying to understand the arrow, postulated an enormous fluctuation, from which we’re still recovering. In that way he could get to low entropy, after which a steady rise in entropy becomes a standard problem in statistical mechanics. The mechanism I just discussed—the change in the dominant force, brought

about by the universe’s expansion, and the transition from a homogeneous state being likely to being unlikely—already explains how entropy could have become small relative to what is dynamically appropriate, without invoking a fluctuation.

Once you have a low entropy state at a particular time in the universe’s history, coupled with the expansion that gives rise to that low entropy state, you have a good case for the induction of the thermodynamic arrow by the cosmological one. As the universe continues to expand, things begin to clump together in galaxies and stars and densities become high enough for nuclear reactions to begin in the cores of stars. Such nuclear reactions represent the slow transition of the stars out of a metastable state (in the first approximation, all matter would like to be iron). These slow transitions keeps us alive and able to maintain our own metastable existence through the negentropy flux from the sun.

In describing the formation of metastable states in the “future” of the change in the dominant forces, I am not falling prey to circularity. It would not be correct to say, “Aha, there is a low entropy state at some time, and the entropy should get bigger in **both** directions away from that minimum (as discussed earlier for the Boltzmann H-theorem).” The dynamics is *not* symmetric about this point in time, precisely because of the expansion.

**Remark:** Another low entropy feature emerging from the dynamics, without having to invoke a fluctuation, is the fact that the material available near the big bang (or big crunch, if things are symmetric) is hydrogen, rather than heavier elements. Thus, once gravity clumps all the low mass nuclides, they can gain energy by combining to form heavier elements.

So there are two pieces to the argument relating the thermodynamic arrow and the cosmological one. First expansion creates a state that is extremely unlikely under the forces dominant at a particular time. Second, given such an unlikely state the entropy increases in the direction of time that we generally call forward. For this second part, the two-time boundary value problem can help. If for the sake of logical consistency you assume a time-symmetric universe, and if you don’t want to impose the arrow of time as a separate postulate, then it is natural to assume that the uniform state of matter also exists sufficiently close to the (future) collapsed state. To know what happens in between, you solve the two-time boundary value problem.

### A. Solving two-time boundary value problems

A number of two-time boundary value problems can be solved. The physical behavior we’re trying to model typically shows good equilibration properties, so that I’ve had my most informative results using mixing transformations like the cat map. (Harmonic oscillators, easy to solve, equilibrate poorly.)

Using the concepts we developed earlier in connection with the study of irreversibility I now pose a two-time boundary value problem for the cat map. I’d like to think of this map as acting on a particles in an ideal gas, say  $N$  points in, as usual,  $\mathbb{I}^2$ . With the coarse graining used in Sec. IF, the two-time boundary condition with low entropy at both ends is posed by demanding that all the points in the gas begin in a single grain, *and also* find their way back into a specific grain at some particular later time, say  $T$ .

In Fig. 14 I show four of the early time steps for a “gas” of 500 points, all of which begin in a single coarse grain. The images do not differ from what is shown in Fig. 4 in any perceptible way (except for the change in grain and in the initial location). Next (Fig. 15) I fast forward a bit showing times 6, 8, 14, and 17. Sometime after 6 you lose any vestige of recognizable pattern. (You’ll see this soon in an entropy plot.) At time step 14 it’s still

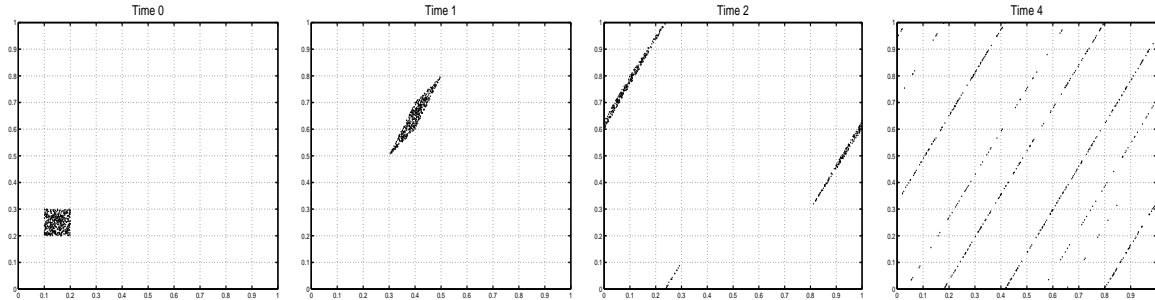


FIG. 14: Four early time steps (0, 1, 2, 4) for 500 points initially placed in a particular coarse grain.

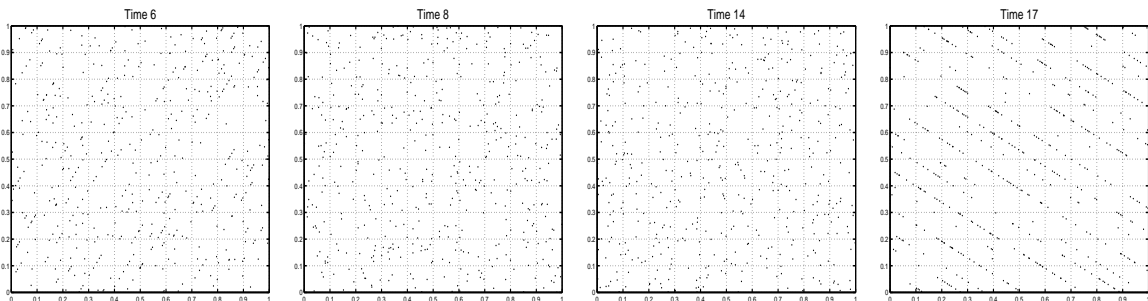


FIG. 15: Time steps 6, 8, 14, and 17.

a big mess. But by 17 you can see something's happening. Continuing in Fig. 16 are times 19, 20, 21, and 22. Now the peculiarity at 17 is explained. The points were on their way to meet a final boundary condition. As a minor technical remark, note that the lines on the return are perpendicular to those on the way out. They're (roughly) in the direction of the eigenvectors of the cat map, and the stretching for the map and its inverse run along different eigenvectors (perpendicular in this case since there are only two, and the matrix is Hermitian).

Two time boundary conditions can be difficult to solve, but because this is an ideal gas, finding the solution I just displayed is easy. All I did was start with (a few more than) 50000 points and keep the 500 of those that landed where I wanted them (after 22 time steps). Removing the others did not affect the motion of those I kept.

The entropy plot for this sequence of images shows complete temporal symmetry. This

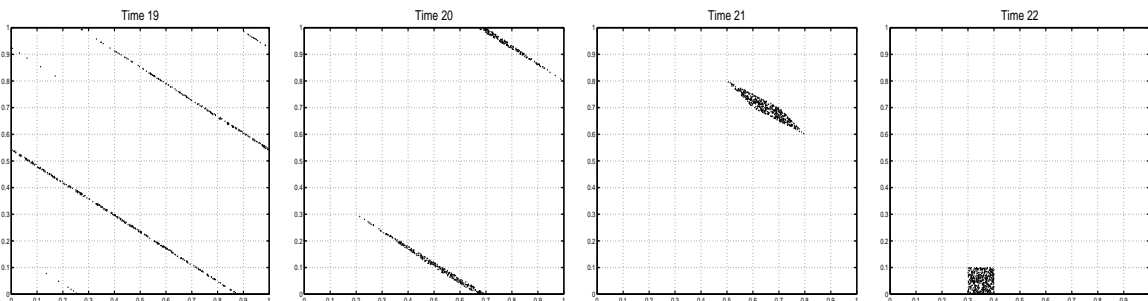


FIG. 16: Time steps 19, 20, 21, and 22.

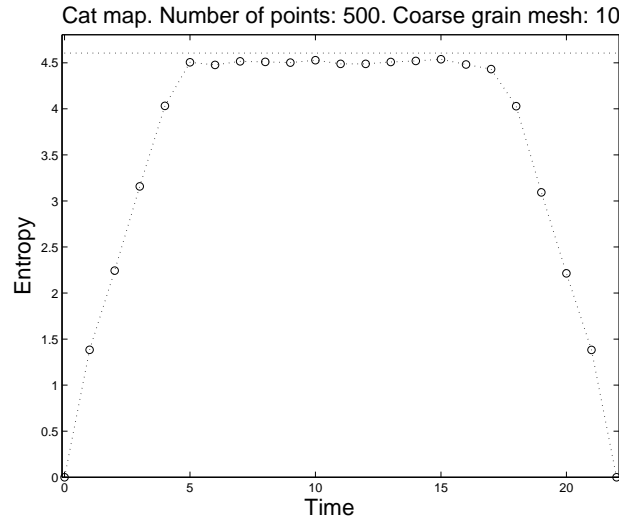


FIG. 17: Entropy as a function of time for the images shown in Figs. 14-16. Compare this dependence to that without future conditioning, shown in Fig. 7.

is shown in Fig. 17. Notice that up to statistical fluctuations it is completely symmetric in time. There is entropy decrease towards the end, but don't worry about the second law! Those who are free from initial condition prejudice are exempt from that law.

A remarkable feature of Fig. 17 is that the initial growth of entropy as a function of time ( $S(t)$ ) is indistinguishable from that *without* conditioning (cf. Fig. 7). I'll show this in a moment in a graph that superimposes  $S(t)$  for several different final conditioning times (“ $T$ ”). This means that even though the points in the initial state ( $t = 0$ ) are very carefully picked, you cannot tell. Or you cannot tell until around time 17, when their otherwise hidden property begins to become manifest. I call this a *cryptic constraint*—the system is subject to some constraint (the final conditioning), but you cannot tell. Every point that has been selected is selected with an accuracy of about  $2.618^{-22} \sim 6.4 \times 10^{-10}$  (which is well within MATLAB's precision). This is because the larger eigenvalue for the cat map is  $\sim 2.618$ , and any error would be magnified by the 22<sup>nd</sup> power of that amount.

It is instructive to ask why this takes place. Here one can use ordinary statistical mechanics reasoning but with a new twist. It's clear from  $S(t)$  that the system forgets its initial conditions in about 5 times steps; for this coarse graining that's the relaxation time. Another way to say this is that wherever the point started, it can be *anywhere* in  $\mathbb{I}^2$  within about 5 time steps. This means that the normal relaxation, despite the future conditioning, can be understood as a kind of forgetting of the final conditioning. Not only can the system get anywhere in 5 time steps, it can get *to* anywhere. So 5 steps before the end there is little indication of its future trajectory.

The manifest symmetry of the graph with respect to the pattern of initial relaxation and final un-relaxation is due to the similarity of the dynamics in both directions, each of them dominated by the large eigenvalue of the cat map matrix. Early on, this symmetry led people to suggest that if Gold was right, sentient beings would *always* see an expanding universe. That is, the arrow of time in the contracting phase, that approaching lower entropy, would be reversed.

I've recently provided further evidence of this by showing that a fundamental notion, macroscopic causality, can be *derived*. Often this is postulated as the basic arrow of time,

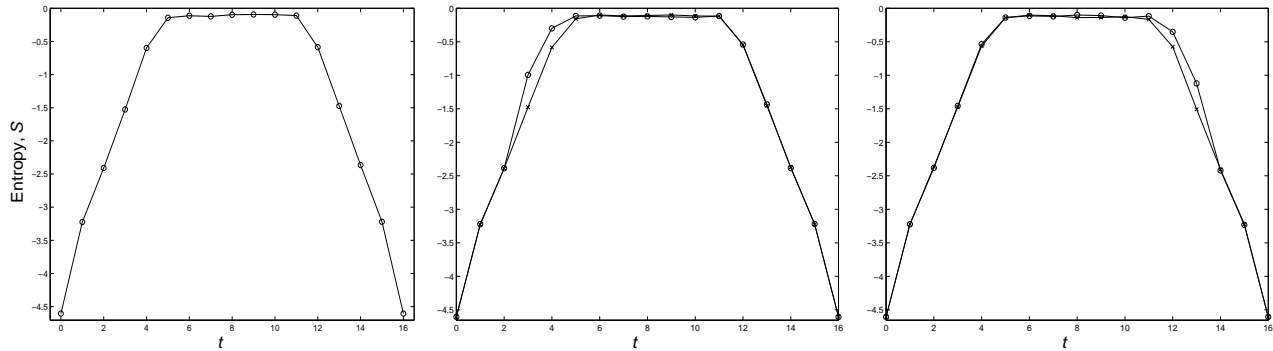


FIG. 18: Causality is an effect: Entropy for the double conditioned cat map. On the left there is no perturbation. The middle figure reproduces the graph on the left, but with the entropy dependence for an early time perturbation superimposed. On the right is again shown the unperturbed graph, this time with the entropy time dependence for a late-time perturbation.

but in the present context, where the arrow itself is something to be derived, causality also is not fundamental. Of course it's not so easy to say what you mean by causality, and once you've said it, you've often not so much given a definition as launched a philosophical debate. (Note, I'm talking about macroscopic causality, not the commutation relation properties that enter in microscopic causality.) The way I chose to interpret this idea was to imagine a double conditioned dynamics in which at one particular time the dynamical rule is changed, but for which the boundary conditions remain the same. That is, you do  $T$  steps of cat map dynamics with a particular pair of boundary conditions. Then, for the same boundary conditions you do  $t_0$  steps of cat map dynamics, one step of something else (e.g., a different measure preserving matrix), and  $T - t_0 - 1$  steps of cat map dynamics. Then you compare the macroscopic behavior for the two. It turns out that if  $t_0$  comes during in interval of non-constant entropy, the change in macroscopic variables is always *later* in the time direction of entropy increase. This is illustrated in Fig. 18, with my standard laboratory technique, the cat map, although there is no problem showing this analytically as well [28]. (Nor is this difficult to understand intuitively, however surprising it may seem on first impression.)

In Fig. 19 I show  $S(t)$  for a variety of conditioning times. As is evident, for times less than ten—twice the relaxation time—there *is* a perceptible difference in macroscopic behavior brought about by the conditioning. The system does not relax normally, in fact it does so more slowly than usual (note too that once  $T > 10$ , you really can't tell what  $T$  is).

This result has physical implications. Determining the cosmology in which we dwell is a central scientific challenge. The idea of an oscillating universe is still on the table, despite the discovery of accelerated expansion. What the ideas inherent in Fig. 19 suggest is a conceptually independent way to get a handle on our cosmology. The first suggestion of this sort was by John Wheeler [29, 30]. He proposed that you go into your lab and check the decay of  $\text{Sa}^{147}$ . If you are careful—and if the time until the big crunch is not much more than 300 Gyr—you'd see not only an exponential with decay rate of about 100 Gyr, but an exponential going the other way, altogether a hyperbolic cosine. Because of the future constraint (back to primordial hydrogen), the sample would be relaxing more slowly than you'd expect it to. The beauty of this idea is that it is a conceptually independent way to study cosmology, based as it is on statistical mechanics, and of course on the assumption that the arrow of time indeed follows the arrow of cosmological geometry.

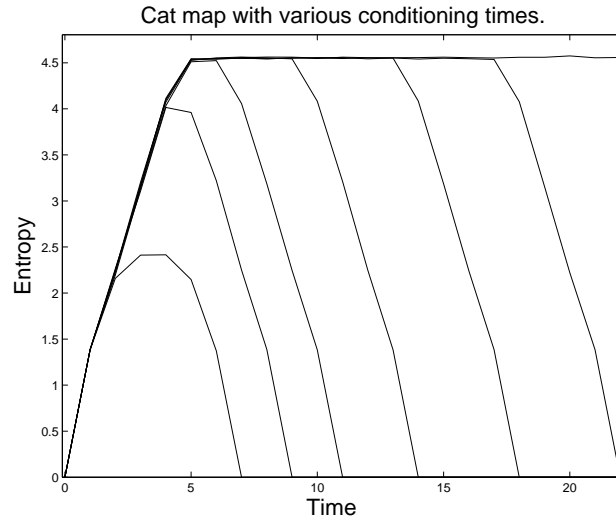


FIG. 19: Entropy for the macroscopically low entropy double conditioned cat map, with various conditioning times.

Unfortunately there are technical difficulties with this idea. By looking at two time conditioning when a small piece is broken off from the main trend [31], I found that a laboratory sample would still decay exponentially, and to see the cosh you'd need to look at cosmological abundance. (Wheeler also retracted his idea [32].) Instead I proposed to look at other dynamical variables with extremely long relaxation times. I believe the best candidate is large scale structure, that is the dynamics of the galaxies themselves. Unfortunately it's not clear what relaxation means in this context and because of a lack of information concerning dark matter and dark energy, it would be difficult to say how quickly something *should* decay. Nevertheless, there's a lot of work on the dynamics of large scale structure, including careful simulations of the motion of many particles (galaxies), including the effects of dark matter, etc. Whether at this point there is any variable whose behavior is well enough understood to say that deviations from observation can be attributed to future conditioning, I doubt. What I believe is worth doing, is trying to find a variable that on the one hand would be sensitive to distant future conditioning, but on the other would be relatively insensitive to the major unknowns that still plague cosmology. Again, I do not know if it's overly optimistic to expect to find anything.

### B. Symmetric behavior of a non-time-reversal invariant dynamical map

In Sec. V we discussed the arrow of time associated with CP violation. At that point I expressed the opinion that I did not believe that arrow had anything to do with our thermodynamic arrow. Besides the smallness of the effect, the asymmetry has no apparent connection to dissipation. There is a different rule for going forward and back, but that in itself does not make your coffee cool.

I will next describe a variant cat map dynamics that does not have time reversal symmetry. Using this asymmetric dynamical law I have done much the same thing that was just done for the cat map itself: two time conditioning with symmetric boundary conditions. The result is shown in Fig. 20. As you can see, despite the non-invariance, this map is as

symmetric or asymmetric as a mapping conventionally possessing time reversal invariance.

*An asymmetric cat map* To do define an *asymmetric* cat map, you have to know what it means for a dynamical law to be time reversal invariant. The way I like to state this is to have only forward time propagation [33]: You have a space  $\Omega$  (for the cat map  $\Omega = \mathbb{I}^2$ ), a time evolution operator  $\phi$  (for some particular time interval [34]), and a putative time-reversal operator,  $T$  (for which  $T^2 = 1$  on  $\Omega$ ). What you need for time reversal invariance is that if you start with any state,  $\omega \in \Omega$ , propagate it forward with  $\phi$ , then apply  $T$ , then propagate with  $\phi$  and finally hit it again with  $T$ , you come back to  $\omega$ . In symbols:  $T\phi T\phi = 1$ , which, for invertible  $\phi$  is equivalent to the more familiar  $T\phi T = \phi^{-1}$ . For classical mechanics  $\Omega$  is phase space and  $T$  the reversal of velocities. In quantum mechanics there is the famous anti-unitary operator involving complex conjugation. For the cat map you can find 2-by-2 matrices that do the job, coupled with the mod 1 operation. For example,

$$T = \begin{pmatrix} 2 & 3 \\ -1 & -2 \end{pmatrix} \quad (87)$$

satisfies  $MTMT = 1$ , with  $M$  the cat-map matrix (Eq. (25)). To produce a time evolution operator that is not time-reversal invariant, we introduce an operator,  $C$  (“Crunch”),

$$C = \begin{pmatrix} 1 & \alpha \\ 0 & 1 \end{pmatrix}, \quad (88)$$

again acting modulo 1, with  $\alpha \in \mathbb{R}$ . Like  $M$ ,  $C$  is measure preserving and 1-to-1, but it has the advantage that it has a continuously variable (non-integer) parameter,  $\alpha$ .

For  $\phi$  we take the product [Crunch]◦[Cat map], with the mod 1 operation performed after each step. (Similarly when  $T$  is applied, one takes mod 1 before and after.) If you ignored the mod 1 operation, you would find that as matrices, not only is  $T^2 = 1$ , but  $\phi * T * \phi * T = 1$ , as well. (This works for  $\phi$  replaced by the cat map alone, which is why the cat map is time-reversal invariant. However, for non-integer  $\alpha$ , the product of these operations is *not* the identity, and  $\phi$  is not a time reversal invariant operation (I have not proved that there is no non-linear transformation that does the job, so this demonstration is not complete).

Using this  $\phi$ , we would like to know whether, on its own, it provides an arrow of time. One way to check is to see whether with two-time boundary conditions it can undo equilibrium just as the time-reversal invariant cat map does. That it has the same symmetry properties is shown in Fig. 20.

### Two-time boundary conditions: Counting states in a cubic centimeter of air

How restrictive is the demand for low entropy two-time boundary conditions? Can all the gas in a room re-collect in a little corner? In classical mechanics if a system is sufficiently chaotic you can get it from anywhere to anywhere (in phase space) on time scales longer than its relaxation time (where that time can be defined with respect to the level of detail with which you have defined “anywhere”)

Quantum mechanics however imposes a scale for the fine graining of phase space, namely  $2\pi\hbar$ . In a  $2n$ -dimensional phase space each region of volume  $(2\pi\hbar)^n$  gives rise to an independent Hilbert space direction (or basis vector). So for quantum mechanics it’s not enough

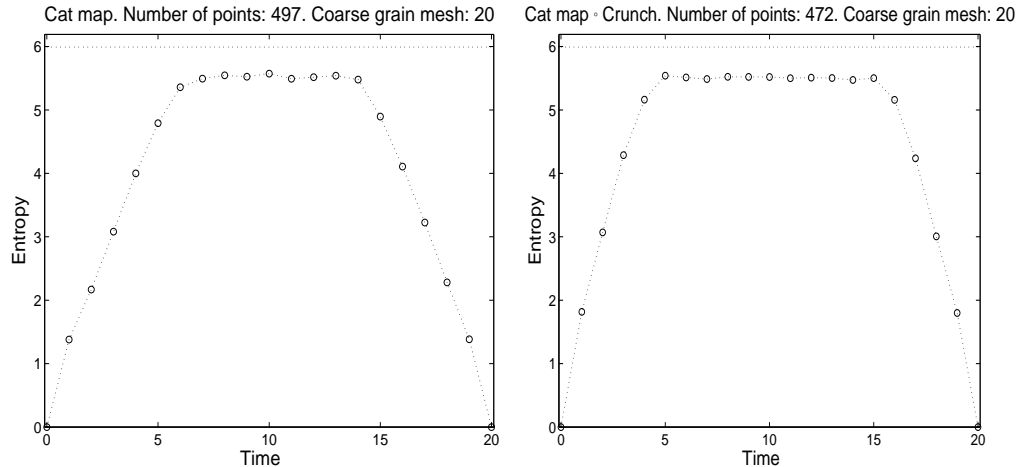


FIG. 20: Time symmetric behavior with two-time boundary conditions. On the left is the entropy for the usual doubly conditioned cat map. On the right the map that is used is the  $\phi$  of Sec. V, which is not invariant under time reversal, although it is non-dissipative and measure preserving.

to find a small region that gets you where you want to go—that region requires a minimum volume before you can satisfy the boundary conditions.

Using the quantum scale I estimate the number of microstates available to the atoms in a cubic centimeter of air. Actually, I'll estimate a smaller number (by a few million powers of ten), the states associated with a monatomic ideal gas of atomic weight 30, at room temperature and at atmospheric pressure. Let  $N_1$  be the number of particles in the given volume. Let  $\bar{p}$  be a characteristic momentum scale for the particles.

**Exercise:** Show that  $N_1 \sim 2.42 \times 10^{19}$ .

**Solution:** Use  $PV = NkT$  with  $V = 10^{-6} \text{ m}^3$ ,  $T = 300 \text{ K}$ ,  $P = 1 \text{ atm}$ .

To estimate the number of microstates we can use standard formulas for the entropy of an ideal gas, taking into account particle identity. From Baierlein [35], p. 103–104, Eq. 5.41:

$$S = kN \ln \left[ \frac{(V/N)}{\lambda_{\text{th}}^3} e^{5/2} \right], \quad \text{with} \quad \lambda_{\text{th}} = \frac{h}{\sqrt{2\pi mkT}}.$$

Letting  $\mathcal{N}$  be the number of states,

$$\mathcal{N} = \exp(S/k) = \exp \left( N \log \left[ \frac{(V/N)}{\lambda_{\text{th}}^3} e^{5/2} \right] \right) \quad (89)$$

In MKS units:  $\lambda_{\text{th}} = 1.8403 \times 10^{-11} \text{ m}$ ,  $V/N = 4.1667 \times 10^{-26} \text{ m}^3$ , and

$$\mathcal{N} = 10^{10^{20.2784}} \quad (90)$$

Now consider various initial and boundary value problems.

Suppose we placed the  $N$  particles in a box 0.25 cm on a side, rather than 1 cm. The ratio the number of states in the 1/4 cm case (“q”, quarter) to that in the 1 cm case (“o”, one) is (from Eq. (89))

$$\frac{\mathcal{N}_q}{\mathcal{N}_o} = \exp(N(\log V_q - \log V_o)) \quad (91)$$



This will be a tremendous reduction of phase space numbers, namely a factor of  $10^{(10^{19.16})}$ . Comparing however to Eq. (90), this is seen to be a minuscule effect, and a number much larger than 1 remains (much, MUCH larger).

Now another scenario: imagine an initial condition in which the particles were in a region 0.5 cm on a side. This cuts the phase space by a factor (“h” for half)

$$\frac{\mathcal{N}_h}{\mathcal{N}_o} = \exp(N(\log V_h - \log V_o)) \quad (92)$$

And now finally we demand a particular final condition: that the particles *also* find themselves in a region 1/2 cm on a side, say 10 minutes later (so this is a two-time boundary condition, requiring confinement in a 1/2 cm box twice). Dynamically there should be classical states that get there, but what we want to check now is whether there are enough to leave a phase space volume containing more than one quantum state. But we already know how much this will cut down phase space: it’s the factor just given in Eq. (92).

To evaluate this, I’ll stop playing with numbers and observe that the ratio in Eq. (91) is just  $\exp(-3N \log 4)$ , while that in Eq. (92) is  $\exp(-3N \log 2)$ . Squaring the second number gives the first! In other words, the cost of the final condition is no worse than a slightly more stringent initial condition. At the macroscopic scale, your pump would have to work a little harder. In this sense, two-time boundary value problems have plenty of solutions. One might still ask why we never see this happen spontaneously. The answer is that although the entropy decrease is not much, it’s still far too large for a spontaneous fluctuation. The odds of such a thing are given by the ratio in Eq. (92).

### VIII. HOW DO YOU DEFINE MACROSCOPIC? ... WITH PRACTICAL CONSEQUENCES

This work is motivated by two problems.

- The second law of thermodynamics: Heat cannot be converted to work, entropy increases (with appropriate delimiters). But the work/heat distinction, or the micro/macro distinction, depend on coarse grains (or their equivalent). What defines coarse grains?
- Nonequilibrium statistical mechanics, and in particular phase transitions
  - For example, metastability is poorly described by equilibrium statistical mechanics, because of the central role played by the (inappropriate for metastability) thermodynamic limit.

When I first began working on this question a friend chastised me: this is *really* a waste of time. (I think the “really” meant to suggest that this was a waste of time even by my usual standards. And this is a friend.)

It turns out however that the issue of defining coarse grains is part of a general problem, both philosophical and practical. One related and currently fashionable issue is the finding of community structure in networks. As discussed in [36], this is also related to how the mind forms conceptual categories, “words” (in a generalized sense, since mice and worms also form such categories (perhaps only on the time scale of the species)).

### A. The role of time

It is my thesis that the primary criterion for the definition of coarse grains is *time*. In studying non-equilibrium statistical mechanics, B. Gaveau and I developed the impression that certain slowly changing quantities that emerged in our formalism played the role of *observables*. This led us to ways to make the definition of coarse grains precise. But before going into this, I want to mention that although our views were formed by the problems we were working on, the perspective, the relation of *macroscopic* to (relatively) slow, was far from original with us. Here's a quote from the introductory pages of Landau and Lifshitz, *Statistical Mechanics* [37],

... Owing to the comparative slowness of chemical reactions, equilibrium as regards the motion of the molecules will be reached ... more rapidly than equilibrium as regards ... the composition of the mixture. This enables us to regard the partial equilibria ... as equilibria at a given ... chemical composition.

The existence of partial equilibria leads to the concept of macroscopic states of a system. Whereas a mechanical microscopic description ... specifies the coordinates and momenta of every particle, ... a macroscopic description is one which specifies the mean values of the physical quantities determining a particular partial equilibrium, ...

### B. The framework: stochastic dynamics

But before implementing grain definition, I need to explain the formalism. I will also take the opportunity to talk about phase transitions, which will be a way of confirming that our “observables” are related to the usual notions of emergent macroscopic concepts.

I introduce basic notions of stochastic dynamics.

The dynamics takes place on a state space  $X$  of cardinality  $N < \infty$ . The motion is a Markov process,  $\xi(t)$ , in  $X$  with transition matrix  $R$ :

$$R_{xy} = \text{“Pr}(x \leftarrow y)\text{”} = \text{Pr}\left(\xi(t+1) = x \mid \xi(t) = y\right). \quad (93)$$

Thus if  $p(x, t)$  is the probability distribution on  $X$  at time  $t$ , the distribution at  $t+1$  is  $p(x, t+1) = \sum_y R_{xy} p(y, t)$ . The matrix  $R$  is assumed to be irreducible, which implies that it has a unique eigenvalue 1 and a strictly positive eigenvector  $p_0$ :

$$\sum_{x \in X} R_{xy} = 1 \quad \text{and} \quad \sum_{y \in X} R_{xy} p_0(y) = p_0(x) > 0 \quad \forall x. \quad (94)$$

By the Perron-Frobenius theorem [38, 39], all eigenvalues of  $R$  lie on or inside the unit circle in the complex plane. These eigenvalues  $\lambda_k$  are ordered by decreasing modulus and increasing phase, if applicable:  $\lambda_0 \equiv 1 \geq |\lambda_1| \geq |\lambda_2| \geq \dots$ . Because  $R$  need not be symmetric (and generally won't be) there are distinct right and left eigenvectors:  $p_k$  and  $A_k$

$$R p_k = \lambda_k p_k, \quad A_k R = \lambda_k A_k, \quad k = 0, 1, \dots. \quad (95)$$

Although  $R$  may not be diagonalizable (and may require a Jordan form), we assume that for the eigenvalues that concern us (those near 1) a spectral decomposition can be used.

The eigenvectors are orthonormal in the following sense:

$$\langle A_k | p_\ell \rangle = \delta_{k\ell}$$

which leaves undetermined a multiplicative factor for each  $k$ . The stationary state  $p_0$  is naturally normalized by  $\sum p_0(x) = 1$ , so that  $A_0(x) = 1 \forall x \in X$ . This leads to several normalization options:

- An “L<sup>1</sup>” perspective:  $\sum |p_k(x)| = 1, \forall k$ .
- Same as for  $A_0$ :  $\max_x |A_k(x)| = 1$
- An “L<sup>2</sup>” perspective, motivated by the fact that when detailed balance holds,  $A_k(x) = p_k(x)/p_0(x)$ . (Detailed balance means  $R_{xy}p_0(y) = R_{yx}p_0(x)$ .) Then the following normalization is natural:  $\sum_x p_0(x)A_k(x)^2 = 1$ .

What will be seen is that the slow (lowest  $k$ ) left eigenvectors of  $R$  are the macroscopic observables. Our take on phase transitions highlights this.

### C. Phase transitions

An idea that goes back to the Onsager solution of the 2-D Ising model is that phase transitions result from eigenvalue degeneracy (for this case the degeneracy is in the eigenvalues of the “transfer matrix” [40]). This was especially advocated by Marc Kac, with a variety of examples. The problem with this dream was that not every phase transition came with a linear operator.

“ $R$ ” provides such an operator with great generality.

We now consider a situation in which there is eigenvalue degeneracy [41, 42]. Assume  $\lambda_1$  is real and let there be a range of times  $t$  such that

$$1 - \lambda_1^t \equiv \epsilon \ll 1 \text{ and } |\lambda_2^t| \ll 1.$$

#### *A basic relation*

The probability distribution at time  $t$ , given that a particle was at  $y$  at time 0, is

$$p_y^t(x, t) = \sum_u R_{xu}^t \delta_{uy} = R_{xy}^t.$$

Since  $A_1$  is orthogonal to the strictly positive  $p_0$ , we deduce that  $A_{1M} \equiv \max_x A_1(x) > 0$  and  $A_{1m} \equiv \min_x A_1(x) < 0$ . Furthermore, since  $|X| < \infty$ , these extreme values are assumed on points of the space,  $x_M$  and  $x_m$  (which may not be unique). The eigenvalue relation for  $A_1$  implies

$$\lambda_1^t A_1(x_M) = \sum_x A_1(x) p_{x_M}^t(x). \quad (96)$$

Divide this by  $A_{1M}$  and subtract it from  $1 = \sum_x p_y^t(x)$ . Do the same for  $A_{1m}$  and  $x_m$ . This yields

$$1 - \lambda_1^t = \sum_x p_{x_M}^t(x) \left(1 - \frac{A_1(x)}{A_{1M}}\right) = \sum_x p_{x_m}^t(x) \left(1 - \frac{A_1(x)}{A_{1m}}\right) \quad (97)$$

These relations provide a quantitative measure for something I noticed long ago about stochastic theories of metastable states: the next to slowest ( $\lambda$  closest to 1) left eigenvector is approximately constant on the phases (including the metastable one).

Eq. (97) explains this feature. Within the sums, all summands are positive. Therefore if the left hand side is small ( $1 - \lambda^t \ll 1$ ) then either  $p_{x_M}^t$  is small or  $[1 - A_1(x)/A_{1M}]$  is. This implies that if it the system is likely to go from  $x_M$  to  $x$  in the time  $t$ , then  $A_1(x)$  is close to  $A_{1M}$ . The same observations hold for  $A_{1m}$  as well. That is, **points that are close dynamically have similar values of  $A_1$** . Note that the relation between dynamical proximity and near values of  $A$  is not symmetric. The  $A_1$  values are close if you can go easily from  $x_M$  to  $x$ , but sometimes you can go from  $x$  to  $x_M$  but not the other way: in that case, the  $A$  values may differ significantly.

This demonstration did not use the smallness of  $\lambda_2$ . Thus for *any* eigenvalue  $\lambda$  and for any time  $t$ , points that are close dynamically on that time scale (i.e., they can reach each other) will have values of the associated  $A$  that are close to each other, where “close” means that the distance is controlled by  $1 - |\lambda|^t$ .

I’ll return to this point later when defining *coarse grains*.

It is natural to define the phase in terms of the “observable”  $A_1$ , which can be considered the name of the phase. For a number  $a$  define

$$I_M(a) = \left\{ x \in X \mid \frac{A_{1M} - A_1(x)}{A_{1M}} < a \right\}$$

$$I_m(a) = \left\{ x \in X \mid \frac{A_{1m} - A_1(x)}{A_{1m}} < a \right\}.$$

For small enough  $a$  these sets will be disjoint. They are the phases.

#### D. Many phases, extension of the basic relation

I use the “max” norm for the  $A$ ’s.

The degeneracy assumption is extended: for some integer  $m$ ,  $\lambda_1, \lambda_2, \dots, \lambda_m$  are real and close to  $\lambda_0 \equiv 1$ . That is, there exists a range of  $t$  such that for some  $\epsilon \ll 1$

$$1 - \lambda_k^t = O(\epsilon), \quad 1 \leq k \leq m. \quad (98)$$

In much of our development we further assume that the  $|\lambda_{m+1}|$  is much smaller than  $\lambda_m$ , that is,

$$|\lambda_{m+1}^t| = O(\eta) \ll 1. \quad (99)$$

As above, for  $x, y \in X$ , define

$$p_y^t(x) = R_{xy}^t. \quad (100)$$

$p_y^t(x)$  is the probability that a system in state  $y$  at time 0 is in  $x$  at time  $t$ .

Consider the following geometric construction in  $\mathbb{R}^m$ : for any  $y \in X$ , form the vector  $\mathbf{A}(y) \equiv (A_1(y), \dots, A_m(y)) \in \mathbb{R}^m$ . This gives a set  $\mathcal{A}$  of  $N$  vectors in  $\mathbb{R}^m$ .

Call this the **observable representation of  $X$** .

Let  $\widehat{\mathcal{A}}$  be the convex hull of  $\mathcal{A}$ .

By the orthogonality relation, the vector  $\mathbf{0} \equiv (0, 0, \dots, 0)$  is in  $\widehat{\mathcal{A}}$ .

$$\sum_{y \in X} p_0(y) \mathbf{A}(y) = \mathbf{0}, \quad (101)$$

so that  $\mathbf{0}$  is a convex combination of the vectors of  $\mathcal{A}$ . As a consequence, one can find  $m + 1$  points  $y_\ell^*$ ,  $1 \leq \ell \leq m + 1$ , such that the vectors

$$\mathbf{E}_\ell \equiv \mathbf{A}(y_\ell^*) \quad (102)$$

are extremal points of  $\widehat{\mathcal{A}}$ , and such that  $\mathbf{0}$  is a convex combination of them. There may be several ways to choose the  $\{y_\ell^*\}$ , but in fact, the  $\{\mathbf{E}_\ell\}$ ,  $1 \leq \ell \leq m + 1$ , are uniquely defined up to a small ambiguity. By the selection of the  $\mathbf{E}_\ell$ , we can find  $\mu_\ell$ ,  $1 \leq \ell \leq m + 1$ , with  $0 \leq \mu_\ell \leq 1$ ,  $\sum_{\ell=1}^{m+1} \mu_\ell = 1$  such that

$$\sum \mu_\ell \mathbf{E}_\ell = \mathbf{0}. \quad (103)$$

For technical reasons we make the following separation hypothesis:

**Hypothesis  $\mathcal{S}$ :** For each  $\ell$  let

$$\phi_\ell = \min_{k=1, \dots, m, k \neq \ell} \|\mathbf{E}_\ell - \mathbf{E}_k\|, \quad (104)$$

and define

$$\Phi \equiv \min_\ell \frac{\phi_\ell}{\|\mathbf{E}_\ell\|}. \quad (105)$$

Then the hypothesis is that the extrema  $y_\ell^*$  can be selected so that (in addition to Eq. (103)) they satisfy the following:

$$1 - \lambda_m^t = \epsilon \ll \Phi \leq O(1). \quad (106)$$

We do not know at this point whether this hypothesis can be dispensed with or whether it is possible to nicely categorize the processes for which it holds. So far we have been able to show it to be superfluous in all cases that we studied in detail.

We next observe that by definition

$$\lambda_k^t A_k(y_\ell^*) = \langle A_k | p_{y_\ell^*}^t \rangle, \quad (107)$$

so that for all  $\ell$

$$(\lambda_1^t A_1(y_\ell^*), \lambda_2^t A_2(y_\ell^*), \dots, \lambda_m^t A_m(y_\ell^*)) = \sum_{y \in X} p_{y_\ell^*}^t(y) \mathbf{A}(y) \quad (108)$$

Because  $0 \leq \lambda_k < 1$ ,  $k \leq m$ , the vector on the left side of Eq. (108) is in the convex set  $\widehat{\mathcal{A}}$ . Using  $\sum_y p_{y_\ell^*}^t(y) = 1$ , we have from Eq. (108)

$$\begin{aligned} \mathbf{E}_\ell - (\lambda_1^t A_1(y_\ell^*), \lambda_2^t A_2(y_\ell^*), \dots, \lambda_m^t A_m(y_\ell^*)) \\ = \sum_y p_{y_\ell^*}^t(y) (\mathbf{E}_\ell - \mathbf{A}(y)) \end{aligned} \quad (109)$$

This relation is an  $m$ -dimensional version of our fundamental relation Eq. (97). Like that equation, the left hand side is small, with distances less than  $1 - \lambda_m^t$ . Unlike Eq. (97), however, the right hand side of Eq. (109) is not manifestly a product of positive quantities. It is possible to overcome this by going to each extremal point and defining a coordinate system that forms a cone coming out of that extremal. In that coordinate system all quantities in Eq. (109) are positive and one can deduce the constancy of the  $A$ 's on each phase.

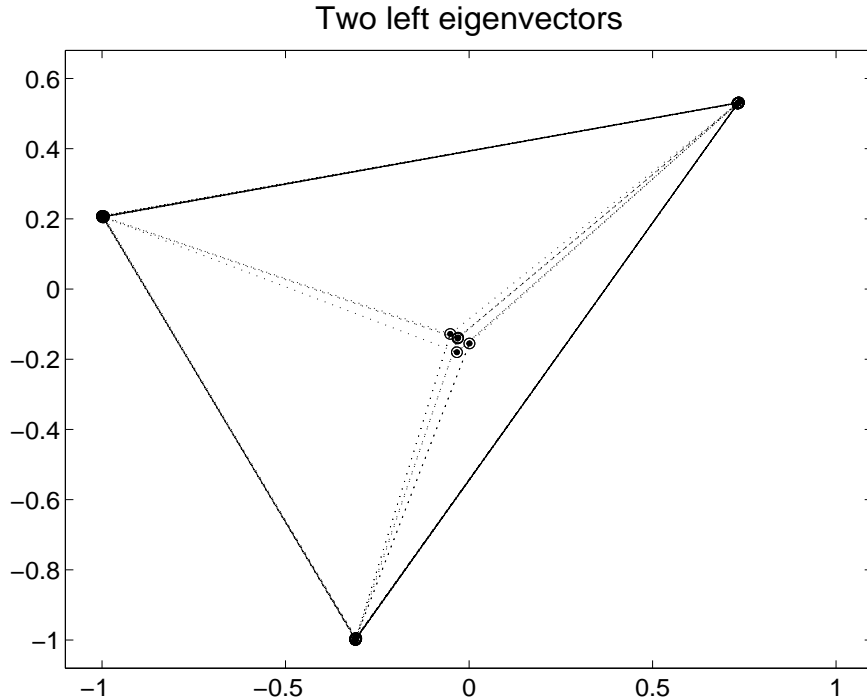


FIG. 21: Plot using the first two left eigenvectors ( $A_1$  and  $A_2$ ) of the transition matrix,  $R$ , for a three-phase system. A circle is placed at each point  $(A_1(x), A_2(x))$  for each of the  $N$  states,  $x$ , in  $X$ . The lines connecting the circles are for visualization. The matrix  $R$  is generated by combining 4 blocks, 3 of which are random matrices, the fourth essentially zero. Then a bit of noise is added throughout, with bigger terms for migration out of the fourth block. Finally the diagonal is adjusted to make the matrix stochastic. This leads to a pair of eigenvalues near one. This plot using the first two eigenvectors shows the extremal points to be clustering in three regions, corresponding to the phases. The points not at the extremals represent the fourth block, all of which head toward one or another phase under the dynamics. For the particular matrix chosen, they are about as likely to end in one phase as another. See [42] for details.

Instead of going through the details, I present an intuitive argument for the reasonableness of the conclusion that the  $A_k$ 's ( $k \leq m$ ) are effectively constant on the phases, and can therefore, collectively, serve as the name of the phase.

The implication of having the  $A$ 's constant on the phases is that all the points from a single phase gather into one small region of  $\mathbb{R}^m$ . That this happens can be seen (for  $m = 2$ ) in Fig. 21. This is a plot of  $A_1(x)$  versus  $A_2(x)$ . The vertices of the triangle that you see are actually composed of many points (shown in more detail in Fig. 22).

Here is a way to understand this bunching. The phases are in a sense dynamically far from one another. If you start in one phase you expect to stay there for a long while before going to any other phase. This means that there is a restricted dynamics within that phase that *nearly* conserves probability. That is, the restricted transition matrix satisfies  $\sum_x R_{xy} \approx 1$ , where the sum over  $x$  is for  $x$  only within the phase and  $y$  is within the phase as well (which implies that the constant vector on the phase is an approximate left eigenvector of eigenvalue near 1). The bunching of points in one phase (as we define it) means that for all points in that phase  $A_k(x)$  has very nearly the same value (for every  $k = 1, \dots, m$ ). Let's see why

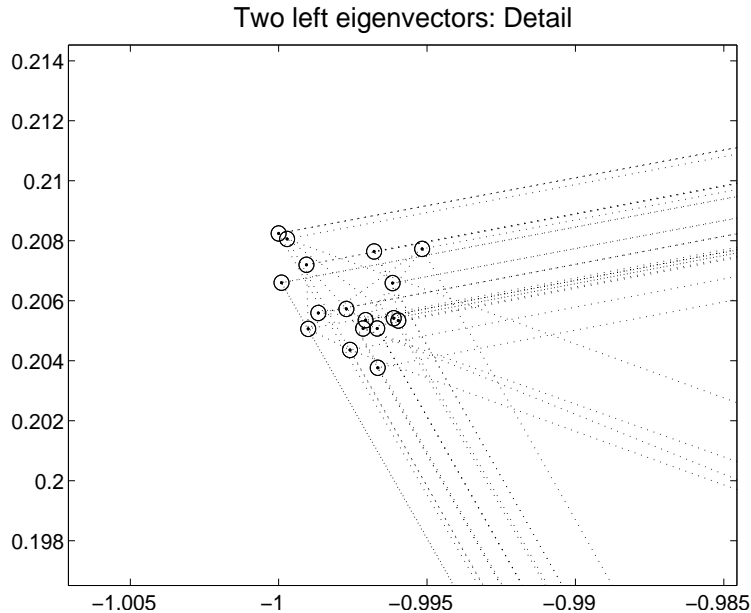


FIG. 22: Detail of the upper left vertex in Fig. 21. In actuality the points in each phase cluster closely together and more than one extremal might, in principle, occur. The precision is limited by the non-negligible magnitudes of the quantities  $1 - \lambda_2$  and  $\lambda_3$ .

that happens. Consider the eigenvalue equation for  $A_k$  applied  $t$  times, where  $t$  is small enough so  $\lambda_k^t$  is still close to one, so close that we will now treat it as unity. Then

$$A_k(y) \approx \sum_x A_k(x) R_{xy}^t. \quad (110)$$

Now suppose  $y$  is reachable with significant probability from  $x$  in time  $t$  (so we would say that  $x$  and  $y$  are in the same phase). Then Eq. (110) can be thought of as an equation within the restricted dynamics. What it says then is that every one of the left eigenvectors  $A_k$ ,  $1 \leq k \leq m$ , provides an eigenvalue 1 for the *restricted dynamics* (up to a small escape rate for transitions between phases). But this left eigenvector is unique and we already have a candidate for it, namely a constant on the phase. It follows that each  $A_k$  ( $1 \leq k \leq m$ ) is proportional to a constant—on each particular phase.

Several other results hold that I will not prove here, but which I will state in the following subsections.

### E. Concentration of the probabilities on the phases

As for the 2-phase case, all points in a phase cluster around the appropriate extremal. In both cases, almost all the weight of the stationary measure ( $p_0$ ) is in the phases. Here it is important that  $|\lambda_{m+1}|$  be small.

## F. Basins of attraction

Our phases include, besides the usual states in a phase, part of the  $O(\eta)$  basins of attraction for those phases. A particular example occurs when the extremal point itself is not in what one usually calls the phase but gets there in  $O(\eta)$ . In this respect the extremal is like the points that are not uniquely identified with any phase and have non-zero probabilities for going to several. The difference between these intermediate points and a not-in-the-phase extremal is that the barycentric coordinates (see below) for an extremal have a single 1, with the other entries zero.

## G. Probability to arrive in a given phase—barycentric coordinates

The points not in the phases lie within the simplex defined by the extremals. Their positions can therefore be written in barycentric coordinates with respect to the simplex. This is a collection of positive numbers adding to 1. These numbers turn out to be the probabilities of starting from a particular point and ending at the corresponding extremal.

This can be seen by observing that for any  $y$  (*not* necessarily in a phase)

$$\lambda_k^t A_k(y) = \sum_x A_k(x) R_{xy}^t = \sum_x A_k(x) p_y^t(x). \quad (111)$$

Now assume that the  $\lambda$ 's are essentially 1 and group the  $x$ -sums into phases, making use of the facts that little of the probability falls outside the phases and that the  $A$ 's are constant on the phases

$$A_k(y) = \sum_\ell \left[ \sum_{x \in X^{(\ell)}} p_y^t(x) \right] A_k(y_\ell^*). \quad (112)$$

The quantity inside the large square brackets is the total probability to go from  $y$  to any point in the phase  $X^{(\ell)}$ . Call this  $q_\ell$ . Eq. (112) holds for each  $k = 1, \dots, m$  and we collect these into a vector equation

$$\mathbf{A}(y) = \sum_\ell q_\ell \mathbf{E}_\ell. \quad (113)$$

This shows that the probability to end in a specific phase is the barycentric coordinate of the associated vector in the observable representation.

An application of this result is a random walk on the landscape shown in Fig. 23. The stationary state is shown in Fig. 24. There are 4 regions of attraction, which we identify as the “phases”. The spectrum of the (225 by 225) generator of the stochastic dynamics is  $[0, \exp(-16.0), \exp(-15.3), \exp(-14.8), \exp(+1.2), \dots]$ , so that it satisfies the conditions for having 4 well-demarcated phases, which in this case represent regions of attraction. Finally in Fig. 25 we show how our methods can be used to calculate the probability that from a given initial condition one will arrive at one or another asymptotic state. This figure is the *observable representation* for this system and forms a tetrahedron (cf. Fig. 26, which was produced in a different way, but is a variation on the theme in that I only display the convex hull). Thus each circle in the graph represents a point on the 15 by 15 lattice and its location in the plot, when expressed in barycentric coordinates with respect to the extremals, gives its probability of reaching a particular phase. In the graph we do not identify the particular circles, but the same computer program that generated the graph can easily provide a table of probabilities for each initial condition.



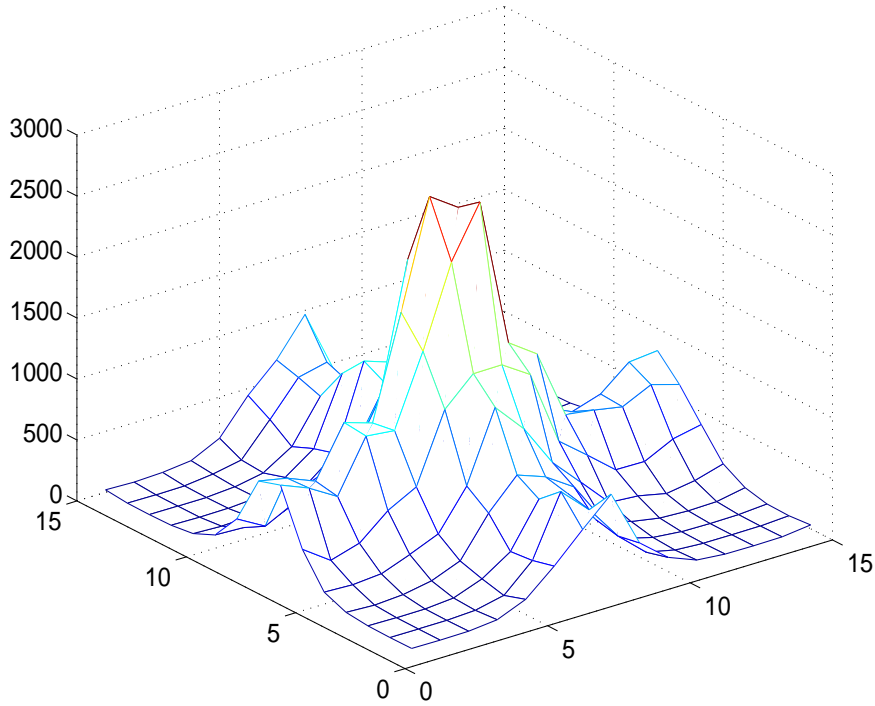


FIG. 23: Landscape for a random walk on a 15 by 15 lattice.

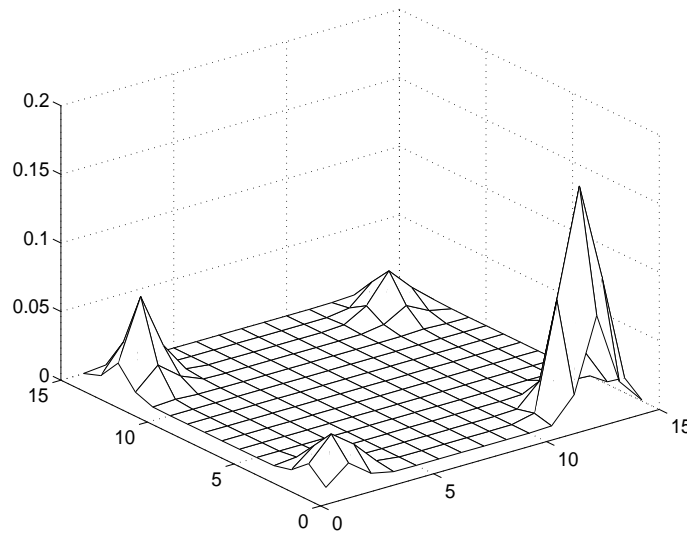


FIG. 24: Stationary distribution for the walk on the landscape shown in Fig. 23.

### H. Hierarchical metastable states

A variation of the “observable representation” which plots the vectors  $\mathbf{A}(x) \equiv (A_1(x), \dots, A_m(x))$ , is to plot  $\mathbf{A}^{(t)}(x) \equiv (A_1(x)\lambda_1^t, \dots, A_m(x)\lambda_m^t)$ . This allows a dynamical image to be viewed. It is particularly informative in the situation where there is a hierarchical relation among the metastable phases (as is supposed to occur in spin glasses). As a function of time phases having the same ancestry merge, finally, as  $t \rightarrow \infty$  coming together in the stationary state. (For some of these applications the eigenvalues drop gradually, but

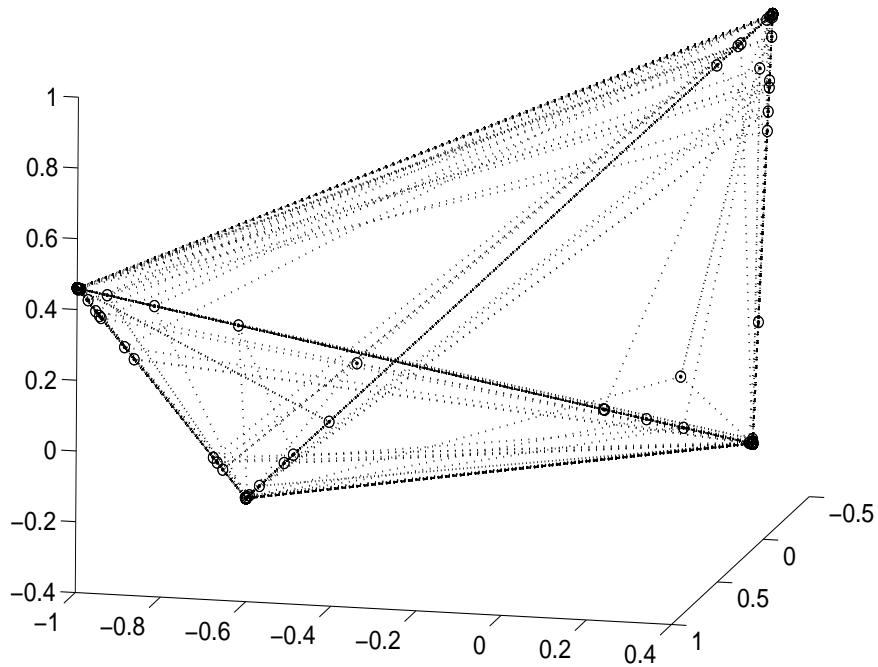


FIG. 25: Observable representation, in  $\mathbb{R}^3$ , of the states for the walk on the landscape shown in Fig. 23. Each circle represents a point on the 15 by 15 lattice and its position within the tetrahedron (when expressed in barycentric coordinates with respect to the extremals) gives the probability of starting at that point and arriving at one or another extremal.

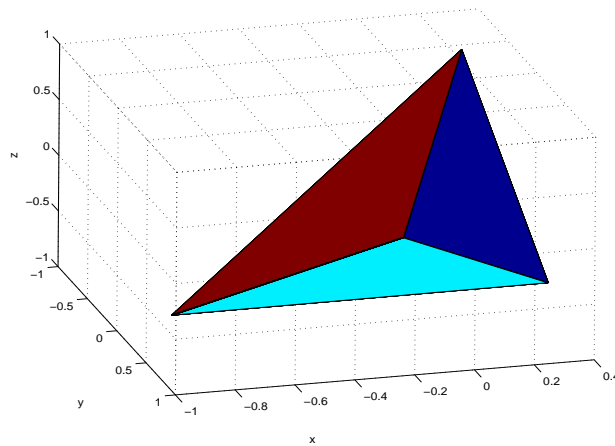


FIG. 26: Convex hull of the set of points  $\mathbf{A}(y)$  for  $y \in X$ . This is for a case of 4 phases and the figure formed in  $\mathbb{R}^3$  is a tetrahedron.

this can still be useful.)

Here is a matrix explicitly possessing hierarchical structure: The overall  $W$  matrix has

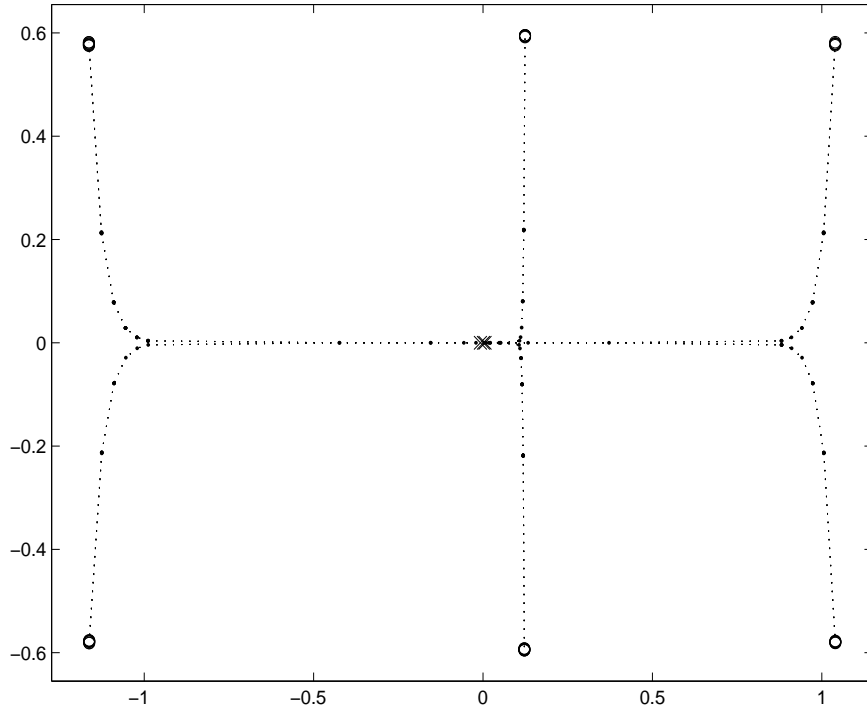


FIG. 27: Phases at successively later times for a hierarchical stochastic matrix. As explained in the text this is a two-dimensional projection of the five dimensional plot of eigenvectors multiplied by eigenvalue to the power  $t$ . On the shortest time scale there are 6 metastable phases (circles); subsequently they merge into three and finally into a single stationary state.

the following form

$$W = \begin{pmatrix} W_1 & \epsilon & \epsilon \\ \epsilon & W_2 & \epsilon \\ \epsilon & \epsilon & W_3 \end{pmatrix}, \quad \text{with each } W_k \text{ of the form } W_k = \begin{pmatrix} w_1 & \delta \\ \delta & w_2 \end{pmatrix}, \quad (114)$$

and  $\epsilon \ll \delta \ll 1$ . ( $\epsilon$  and  $\delta$  are generic small matrices and are not all the same.) There are 6 phases; on a medium time scale three pairs decay into a common branch, subsequent to which the three branches merge into a single trunk. Since we cannot image the 5-dimensional structure, we take the projection of this motion (as a function of time) on a particular plane. This is shown in Fig. 27, where the circles represent the original phases and the “ $\times$ ” is the final state,  $(0, 0)$ .

## IX. COARSE GRAINS

Our nonequilibrium formalism provides a framework within which to implement the physical idea that time is related to “macroscopic.” The first step is to define a metric on  $X$  based entirely on dynamics.

Two points  $x, y \in X$  are more like one another if their values of  $A_k$  are closer; the larger  $|\lambda_k|$ , the more important that closeness is. Define the following distance on  $X$

$$d_T(x, y) \equiv \sum_{k \geq 1} |A_k(x) - A_k(y)| |\lambda_k|^T. \quad (115)$$

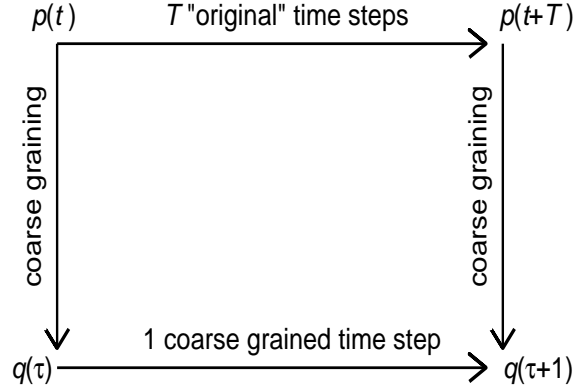


FIG. 28: For grain diameter  $\leq a$ , the two routes from  $p(t)$  to  $q(\tau + 1)$  differ by at most  $a$ .

The parameter  $T$  in Eq. (115) reflects temporal precision.

A coarse graining is a disjoint partition of  $X$ . The grains will be called  $\Delta$  or  $\Delta_k$  and satisfy  $\cup_k \Delta_k = X$  and  $\Delta_k \cap \Delta_{k'} = \emptyset$  for  $k \neq k'$ . Points within a grain are designated  $x$  or  $x_\Delta \in \Delta \subset X$ .

Coarse graining of probabilities is straightforward:

$$\tilde{p}(\Delta) = \sum_{x \in \Delta} p(x). \quad (116)$$

This rule applies to row indices or kets. For column indices, or bras, there is a weighting (necessary to make the coarse grained stationary state a stationary state of the coarse grained evolution). For a grain  $\Delta$ , and for  $y \in \Delta$ , let

$$w_\Delta(y) \equiv \frac{p_0(y)}{\tilde{p}_0(\Delta)} = \frac{p_0(y)}{\sum_{y' \in \Delta} p_0(y')}, \quad (117)$$

where  $p_0$  is the stationary state of  $R$ . Thus the observable  $A(x)$  is coarse grained to become  $\tilde{A}(\Delta) \equiv \sum_{y \in \Delta} w_\Delta(y) A(y)$ .

With respect to the stochastic dynamics itself, an additional time-smearing is applied to the stochastic matrix  $R$ . For fixed  $T$  define

$$\widetilde{R^T}_{\Delta\Delta'} \equiv \sum_{x \in \Delta} \sum_{y \in \Delta'} (R^T)_{xy} w_{\Delta'}(y), \quad (118)$$

The “ $T$ ” in  $\widetilde{R^T}$  is not a power, but a reminder that the original  $R$  (on the right-hand side of Eq. (118)) has been taken to the power  $T$ , i.e., that the time has been coarse grained. It follows immediately that  $\tilde{p}_0$  is the eigenvalue-1 right eigenvector of  $\widetilde{R^T}$ .

An important property for a coarse graining is that it be close to “commuting,” namely  $\widetilde{R^T} \tilde{p}$  be close to  $\widetilde{R^T p}$ . In [36] we show that if  $\forall u, [d_T(x, y) < a, \forall x, y \in u]$ , then  $|\widetilde{R^T} \tilde{p} - \widetilde{R^T p}| < a$ . This is illustrated in Fig. 28.

#### *Random walk: time recovers space*

Consider diffusion on a ring of  $N$  sites,  $1, \dots, N$  (with periodic boundary conditions). Define the matrix  $B$  as translation by one step counterclockwise (e.g.,  $B_{12} = 1$ ). The

transition matrix is  $R \equiv \alpha \mathbf{1} + (1 - \alpha) (B + B^\top) / 2$ , with  $0 \leq \alpha < 1$  and  $\top$  indicating matrix transpose. The eigenvalues and eigenvectors are

$$\begin{aligned} \lambda_k &= \alpha + (1 - \alpha) \cos \phi_k, & \phi_k &= \frac{2\pi k}{N}, \quad k = 0, 1, \dots, \left\lfloor \frac{N}{2} \right\rfloor \\ u_k(x) &= N^{-1/2} \cos(x\phi_k), & v_k(x) &= N^{-1/2} \sin(x\phi_k) \\ & & k &= 1, 2, \dots, \left\lfloor \frac{N}{2} \right\rfloor, \quad x = 1, 2, \dots, N \end{aligned}$$

where  $[w]$  is the integer part of  $w$ . For even  $N$ ,  $v_k$  is zero for  $k = 0, N/2$ , and for odd  $N$  it vanishes for  $k = 0$ . We use a variation of the distance function defined above: Let  $\tilde{d}(x, y)$  be the square root of

$$\tilde{d}_t^2(x, y) \equiv \sum_{k \geq 1} \frac{(A_k(x) - A_k(y))^2}{1 - |\lambda_k|^t} \quad (119)$$

Using  $t = 1$ , with not too much manipulation one obtains

$$\tilde{d}_1^2(x, y) = \frac{2}{(1 - \alpha)} \frac{1}{N} \sum_{k=1}^{\lfloor N/2 \rfloor} \frac{\sin^2((x - y)\phi_k/2)}{\sin^2(\phi_k/2)} \quad (120)$$

For large  $N$  this becomes

$$\tilde{d}_1^2(x, y) = \frac{1}{(1 - \alpha)} \frac{2}{\pi} \int_0^{\pi/2} \frac{\sin^2((x - y)\theta)}{\sin^2 \theta} = \frac{|x - y|}{(1 - \alpha)} \quad (121)$$

( $x$  and  $y$  are integers). Thus the dynamical distance reconstructs configuration space.

## A. Making grains

Defining a distance function provides a basis for the selection of coarse grains. This is not the end of the story, since you need a cluster algorithm to partition the space. We used a Monte Carlo scheme, with an annealing protocol, that minimizes the sum of the internal distances within each grain. The results for various examples are now shown.

### 1. Dynamic Ferromagnet

The one-dimensional Ising model has the energy function

$$H = -J \sum_{k=1}^N \sigma_k \sigma_{k+1} - B \sum_{k=1}^N \sigma_k, \quad (122)$$

where  $\sigma_k$  is a  $\pm 1$ -valued spin at site  $\#k$ ,  $J > 0$  is the spin-spin coupling constant, and  $B$  the external magnetic field. Index addition in Eq. (122) is modulo  $N$ . Many stochastic-dynamical schemes are consistent with this energy and we use the following. A site,  $k'$ , is randomly selected and the change in energy that would occur *if* that spin were flipped,

TABLE II: Magnetization of the states within a grain. Each row corresponds to a grain and lists the magnetization of each state in that grain. Note that within each grain the values of magnetization are almost all equal, demonstrating that with a dynamically defined distance the order parameter, magnetization, emerges naturally.

High temperature, $T = 2.25$	Low temperature, $T = 0.8$
2 2 2 2	2 0 0 2
-2 -2 -2 -2	-2 -2 -2
0 0 0 0	-2 -2 -2 -2
0 0 0 0	4 2 2 2 2
0 0 0 0	0 0 0 0 0
2 2 0 0	-2 -4 -2 -4 -4 -6
0 0 -2 -2	6 4 4 4 4 4
0 0 0 0	-4 -2 -2 -4 -4
6 4 4 4 4 4	2 2 2 2
-4 -4 -4 -4 -4 -6	0 0 0 0
2 2 2 2 2	0 0 0 0 0
-2 -2 -2 -2 -4	-2 -2 -2 -2
4 2 2 2 2	0 0 0 0
-2 -2 -2 -2 -2	2 2 2 2 2

$\Delta E \equiv H(\sigma_{k'} \rightarrow -\sigma_{k'}) - H(\text{original})$ , is calculated. The flip is then implemented with probability  $\alpha \exp(-\Delta E/2T)$ , where  $T$  is a temperature parameter and  $\alpha$  a global constant chosen small enough so that the diagonal components of the transition matrix are non-negative. A second allowed process is a double flip by opposite pointing neighbor spins. The probability that this occurs is governed by an additional controllable parameter. In this model there is no breakdown in analyticity, but for  $T \leq 1$  and  $B = 0$  the system will spend most of the time strongly polarized. ( $p_0$  as a function of magnetization is strongly peaked at symmetric nonzero values.) Above  $T = 2$  the system is found mostly at small magnetization values.

The transition matrix,  $R$  is  $2^N \times 2^N$ , but it is sparse. As expected, for  $T \leq 1$  there is near degeneracy of  $1 \equiv \lambda_0$  and  $\lambda_1$ . (As  $T$  increases above 1, the gap,  $1 - \lambda_1$ , rapidly increases.)

For this example we calculated a set of grains. What emerged was a sorting according to magnetization, both above and below the transition. The magnetization of the states *within* each grain is given in Table II. For the ferromagnet the dynamical distance has revealed an “emergent” quantity, the magnetization.

## 2. Words (organization of concepts) and the brain

For this example and the following one, the matrix  $R$  does not reflect a manifest dynamics, but takes some measure of the connectivity of a set and builds from it a stochastic matrix.

The “states” for the following analysis are elementary units of the brain. They might be neurons, columns or some other structures. The unit can receive input from several sources and can discern the origin of each signal. One idea [43] of how the brain generates “words” is that this unit learns to recognize patterns of stimulation, which we will take to be temporal, sequential patterns, the list of successive signal sources. Our stochastic

matrix,  $R$ , is associated with a single unit and focuses on its input. Take  $R$  to provide the following information: it gives the probability that a signal comes from source  $x$ , given that the preceding signal was from source  $y$ .  $X$  is the list of sources,  $1, \dots, N$ .

Suppose the firing sequence 1-2-3 is common. According to [43], this is eventually recognized by the unit and becomes a generalized “word.” This not only categorizes data but can help in the interpretation of ambiguous signals. The biochemical mechanisms by which this recording takes place are mostly unknown.

We now show how such a pattern, reflected in the  $R$  associated with this unit, will create a coarse grain. For simplicity suppose that the only marked pattern that is present is  $1 \rightarrow 2 \rightarrow 3 \rightarrow 1 \dots$ . This means that  $R$  has matrix elements  $R_{13} \approx R_{32} \approx R_{21} \approx 1$ , with the other matrix elements smaller and essentially random.

Numerically we found that distances between 1, 2 and 3 were much smaller than those between the others. Under coarse graining protocols the coarse grains that gave minimal internal distances were those in which 1, 2, and 3 were put together. This property persisted for the non-cyclic process  $1 \rightarrow 2 \rightarrow 3$ .

A surprising feature was the role of noise. With zero noise,  $R$  (for the cyclic pattern) is simply the permutation on 3 objects, and is the identity on all other states (not reducible, but never mind). The eigenvalues associated with the permutation are the cube roots of unity, all of magnitude one. The left eigenvector for the root 1 is no problem as it is unity on all 3 of the important sources, hence yields small distance. For the other two root-of-unity eigenvalues, the norm of the eigenvalue is *also* unity, but the (left) eigenvectors have large differences (they too have components that are roots of unity). So the distance between these states would not be small (which we want it to be).

With noise this situation clears up in an interesting way. A bit of noise affecting all sources first makes  $R$  irreducible, but also brings in all but one of the eigenvalues from the unit circle. The eigenvalues *most* affected by the noise are non-unit roots of unity, which have associated left-eigenvectors for which the participants in the cycle have quite different values. *For all other* left eigenvectors, the values taken on the cycle-participating sources are nearly the same.

### 3. The karate club

This is an example studied as a test of the formation of community structure in a small network. Many years ago, Zachary [44] studied the breakup of a particular karate club into two factions. He also managed to define a level of relationship between individuals, and since then generations of those studying community structure in networks have been using these relationships to predict the details of the split, with more or less success.

Starting from the adjacency matrix  $B_{xy}$ , describing which individuals are closest, the stochastic matrix can be defined as  $R_{xy} = B_{xy}/n_y$  (no summation) where  $n_y = \sum_x B_{xy}$ . This immediately leads to the spectral properties of  $R$  and to the derived metric. A coarse graining into two grains using this algorithm either gives the correct split, or misses by a single individual, the same individual that in other methods also appears to behave in a maverick fashion. (In our coarse graining more than one outcome is possible because of the Monte Carlo method and also because different temporal parameters were used as well as slight variations in the metric itself.)

## B. Reflections

The effectiveness of the coarse graining scheme shows that in a sense time reconstructs space. And not only space, but order parameters as well (I omitted another example we studied, heat transmission, in which energy density emerged.) Others have found similar results. For example, Halliwell [45, 46], working at the quantum level, shows that the quantities most effective at decohering, hence having the best chance to make it to the classical level, are those that are slowest to change, for example, local energy density.

It's amusing that some people have eliminated time from dynamics [47]. We go in the other direction, showing the pre-eminence of time in the creation of macroscopic perceptions.

## X. QUANTUM MEASUREMENT THEORY

In this section I will mostly talk about my own work in quantum measurement theory. However, irrespective of the merits of my proposal, there are several points of general interest that anyone who worries about this question should come away with. First, there is the realization that quantum measurement theory is a problem in statistical mechanics. This, perhaps surprising, assertion reflects the view that statistical mechanics studies the interface between the macroscopic and microscopic. You derive equations of state by from particle dynamics—for lots of particles. The “problem” of quantum measurement theory is exactly the reconciling of micro and macro, the fact that experiments appear to have a definite outcome, while pure unitary time evolution gives rise to superpositions of living and dead cats. *Except for this apparent contradiction, there is no measurement problem*, and, given the success of other quantum predictions it would never enter anyone's mind to postulate funny mechanisms for “collapse of the wave function” or other nonsense (if I may allow my opinions to show). A second general message from the forthcoming presentation is that “measurement” should not be considered black magic. The separation of the world into system and apparatus, with apparatus treated more or less classically, could be maintained in 1930, when Bohr and Einstein famously debated these issues. But today, with fully quantum treatments of phenomena reaching the mesoscopic level, there is no excuse for not treating the apparatus as another quantum system, albeit a big one. Moreover, its “bigness” is reflected not only in its having a large total mass (as some simple-minded models of apparatus have done), but also in its having many degrees of freedom. A bona fide measurement amplifies and registers. That is, it promotes a microscopic event to one that we sense at a meso- or macroscopic level and it does it in such a way that something irreversible happens. These processes again require many degrees of freedom, so that we have been brought back to statistical mechanics, fleshing out my earlier contention. As part of the presentation of my work, and as a way to check it, I will give an example below of a fully quantum apparatus. If you happen to have your own ideas on quantum measurement, you should also be able to check them in this way. The apparatus model is one I developed with Gaveau, since at that point (about 1990) I found nothing suitable, or even vaguely approaching the realistic, on which to test my ideas. Other models have since been proposed, and besides those in my book [1] I mention that of Allahverdyan et al. [48, 49].

For the purposes of the ideas that I'm about to present, the most important message from the previous lectures is that you should *abandon your initial conditions prejudice*. The “natural” use macroscopic of initial conditions is equivalent to the thermodynamic arrow



of time. If that arrow arises from something else, for example cosmology, then just as one contemplates various cosmologies that may govern a universe, so one can contemplate different arrows of time.

As remarked, the puzzle of quantum measurement theory is the appearance of definite macroscopic states as the result of measurements. The explanation I am about to offer takes quantum mechanics as inviolate and the wave function as physical. The revolutionary steps are in statistical mechanics, challenging the foundations of that theory in areas where it does not have experimental support. And the main challenge will be that when you see a macroscopic state, you do *not* assume that all microstates consistent with the macrostate are equally likely. In particular, you may find *cryptic constraints*, as we did for the two-ended low-entropy ideal gas, evolving under the cat map. Thus statistical mechanics enters profoundly and in a way that goes beyond the treatment of a many-degree-of-freedom apparatus. That's the end of the story; now let me start from the beginning.

### A. Precis of the idea: the schizoid cat

Here's an example of the quantum measurement problem: you have a two-state system described by a wave function  $\phi$ . At the beginning of the experiment,  $\phi = \alpha_1\phi_1 + \alpha_2\phi_2$ . ( $\phi_k$  is assumed normalized and  $\sum |\alpha_k|^2 = 1$ .) The wave function of the apparatus measuring this system is  $\Omega$ . Initially the total wave function is

$$\Psi_i = (\alpha_1\phi_1 + \alpha_2\phi_2)\Omega. \quad (123)$$

The total Hamiltonian is  $H$ , and after a measurement taking time  $t$  the wave function becomes

$$\Psi_f = \exp(-iHt/\hbar) [(\alpha_1\phi_1 + \alpha_2\phi_2)\Omega] = \alpha_1\phi_1\Omega_1 + \alpha_2\phi_2\Omega_2, \quad (124)$$

where  $\Omega_k$  are macroscopically distinct states, meaning that no practical apparatus (having operators  $Q$ ) can get a nonzero value for  $\langle \Omega_1 | Q | \Omega_2 \rangle$  [50]. This is a “Schrödinger cat,” and is never observed, although as I indicated it *cannot* be observed.

At this point there are numerous strategies to justify the statement that one of these two states is randomly selected with probability  $|\alpha_k|^2$  and the other ... disappears, or is irrelevant, or is part of another “world”, or ...

My contention is that among the myriad of macroscopically indistinguishable microstates ( $\Omega$ 's) there are those for which—under pure unitary time evolution—the final state is only *one* of the two possible macrostates. For example, I postulate that there is a state  $\tilde{\Omega}'$  for which

$$\Psi_f = \exp(-iHt/\hbar) [(\alpha_1\phi_1 + \alpha_2\phi_2)\tilde{\Omega}'] = \phi_1\Omega'_1 \quad (125)$$

(in fact there are many of them). Correspondingly there are other initial apparatus states  $\tilde{\Omega}''$  for which

$$\Psi_f = \exp(-iHt/\hbar) [(\alpha_1\phi_1 + \alpha_2\phi_2)\tilde{\Omega}'] = \phi_2\Omega''_2. \quad (126)$$

Thus when you do an experiment and get outcome #1, it means that the apparatus, prior to the experiment, was in a special state of type  $\Omega'$ . Similarly for #2. The superposition principle still works, and if your initial apparatus state were  $\Omega' + \Omega''$ , you would get a Schrödinger-cat state. It's just that this superposition does not appear (as an initial condition) in Nature. A little thought about the implications of this idea shows there is (at this stage) no role for probability and that there must be a single wave function for the universe.

The assertions of the last paragraph about the existence of special states are statements about quantum mechanics, solutions of Schrödinger’s equation for big systems. They are difficult assertions, especially since when I first began thinking about these things I found no reasonable quantum models of apparatus on which to test them. But I’ll show you examples where you can find such states. Let’s move on to the *really* difficult part.

States that lead to only one macroscopic outcome, as in Eq. (125) or Eq. (126), I call “special” states. These equations tell you that if you had as your initial conditions in some experiment a special state (I’ll dispense with the quotation marks) then you have no cat problem. Good, but why should you have this sort of initial condition?

Our experience with two-time boundary value problems suggests that a final condition can select, in a precise—meaning microscopic—way, particular initial conditions. Thus in our doubly-conditioned cat map we could get a Poincaré recurrence in 22 time steps, where ordinary probability arguments would suggest you’d have to wait  $G^N$  time steps, with  $G$  the number of coarse grains and  $N$  the number of particles (which in our examples came to  $\approx 10^{1000}$ ). In thinking about the early universe it is reasonable to suppose it to have been in a pure state in which distant objects were not entangled. I will now assume that we live in a roughly time-symmetric universe, starting at a big bang, ending in a big crunch. If we assume that the arrow of time is a consequence of the geometry then the for the late universe (in our time reckoning) one should have a similar state—and in particular one in which distant objects are not entangled. This will be problematic if unitary evolution is continually generating Schrödinger cats. One way to have distant objects unentangled “late” in the life of the universe would be to maintain purity and localization throughout. The alternative would be to have macroscopically different states (the wave functions that went off to become other “worlds” in the many-worlds sense) recombine coherently. Having distant objects unentangled as the big crunch approaches is unlikely, but the *least unlikely* way to accomplish this is to avoid macroscopic entanglement at every stage. Thus maintaining “purity” (of the wave function) throughout means that at each juncture where a “cat” could form, it doesn’t, namely the state of the apparatus and system is special, in the sense of Eqs. (125-126).

I point out that the use of special states to solve the measurement problem, and the explanation just offered for the actual occurrence of special states in Nature, are conceptually independent. If special states really are the answer, it would not be the first time someone has found a solution without understanding why it works. When Boltzmann (*l’havdil*) postulated an early low-entropy state to solve the arrow of time puzzle he attributed it to a fluctuation. He could have no idea of the developments in cosmology that would provide a much better explanation for the universe to arrive at an extremely unlikely state (see Secs. VII and VII 1). In other words, there could be an entirely different reason for the cryptic constraint that selects special states, but I don’t have the imagination to think of it.

It also must be emphasized that the idea presented here is a physical one, not an alternative “interpretation” (and in particular it is *not* a hidden variable theory). There are particular microscopic states that occur in Nature and there are those that don’t. To show that this idea describes, or may describe, the physical world several tasks must be addressed:

- Do there exist special states? Sec. X B. As indicated, one can examine particular models of apparatus to ascertain whether they have a sufficiently rich class of microstates to give the single-outcome time evolution demanded.
- How do you recover the Born probabilities? Sec. X C. Probability will enter as it

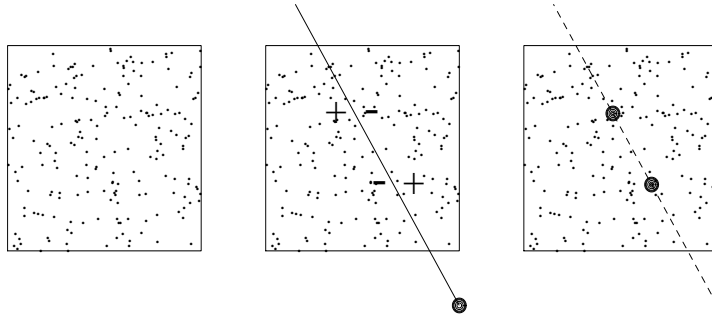


FIG. 29: Three stages in the operation of a cloud chamber. After expansion (left figure) the vapor is supercooled, but in the absence of disturbance would not form liquid for a long time. The middle figure shows a particle traversing the chamber, ionizing two atoms in its wake. On the right, the disturbing particle has long passed and droplets have formed at the sites of the ionization. One deduces the path of the particle (dashed line) from the droplets.

does in classical mechanics. When there are many special states, you cannot know which microstate the apparatus actually is in (same as when you flip a coin). The probabilities are then the measures (relative dimensions of the Hilbert subspaces) of the states associated with each outcome.

- Can experiment determine if this theory is correct? Sec. XD. Something may be physically true even if you can't, today, do an experiment to prove it (string theorists will be sympathetic to this assertion). Nevertheless, I will suggest tests that can be made of this theory.

## B. Existence of special states

### *A schizoid cloud chamber*

I will exhibit special states in a model of a cloud chamber. First I'll review a bit about cloud chambers, then offer a fully quantum model and show how a measurement would create a superposition of macroscopically distinct states. Then I'll identify the special states within the model.

A cloud chamber (see Fig. 29) detects the path of a particle, say a cosmic ray, by forming small droplets of liquid within a vapor (the “cloud”) along the path [51]. Like all apparatus, it functions as an amplifier to promote the microscopic to the macroscopic. This is accomplished by putting the vapor into a metastable state. Within the chamber, besides the material of the vapor, there is a non-condensable gas. By expanding the chamber, the temperature is rapidly lowered. At the new pressure and temperature the vapor should be liquid, but this does not happen instantly because the system first must form a sufficiently large droplet (larger than a “critical droplet” [52]). The passage of the cosmic ray helps create such droplets. After that they grow rapidly, although in practice this growth only takes place long enough for the whole thing to be photographed. Then the chamber is compressed, warming it and reestablishing the vapor.

In any small volume of supercooled gas through which the charged particle passes, there may or may not be an ionization event due to that passage. The scattering and (possible)

ionization are describable by unitary time evolution within a quantum mechanical framework. If one were to calculate the results of the scattering there would be two significant components to the wave function, one in which the atom was ionized and one in which it was not. These are the Schrödinger cat states. The relative norms of these components give the probabilities of droplet formation and non-formation within that volume. (This way of generating a superposition of macroscopically different states is not quite the same as presented in Eq. (124), where the initial wave function already has a superposition structure. The principle however is the same. Below, in Sec. X C, I'll discuss states having the form of Eq. (123).)

I will next make a model of this system, not terribly realistic, but nevertheless possessing the essential properties of the cloud chamber: metastability, amplification and irreversibility. So it will behave like an apparatus. Then I will display its special states.

**Remark:** In the next several pages I present a number of detailed calculations, and it will be easy to get bogged down in details and lose sight of the point of all this arithmetic. Let me here outline the steps.

- Defining the Hamiltonian
- Qualitative description of the operation of the detector
- Calculations supporting the qualitative description
  - Spontaneous spin flip: false positives and ground state stability
  - Scattering-induced spin flips
- Resumé of the detector operation
- Special states. This is the point of the story! After slogging through the detailed operation of the detector, you will see that indeed, although it usually (i.e., for most initial conditions) gives a “cat” state, for some particular initial conditions it gives only a single outcome, one or the other *definite* state of the “cat.”

### *Defining the Hamiltonian*

For the metastable system I use a three-dimensional array of quantum spins. They interact with ferromagnetic short range forces. In the presence of a magnetic field they align along that field. On the left in Fig. 30 I show a two-dimensional slice of an array in which a magnetic field points up. For convenience I assume the temperature is so low that even for a fairly large array no spin is reversed [53]. To produce a metastable state one makes a sudden reversal of the magnetic field. The upward pointing spins are no longer the lowest state of the system, but their transition to the stable state is slow and the system is metastable. If the field is strong the “critical droplet” size can be 1, so that with a single spin flip it will be energetically favorable for the region of overturned spins to grow.

Here is how the array functions as a detector. The particle to be detected, called  $\chi$ , interacts with the spins in such a way that it tends to flip them. When  $\chi$  traverses the array it can cause a spin to reverse. See the right hand image of Fig. 30. The neighbors of that spin are now in a more favorable situation for their own flips. With no spins reversed, six neighbors oppose a flip (on a cubic lattice). But after a reversal, a neighboring spin has only

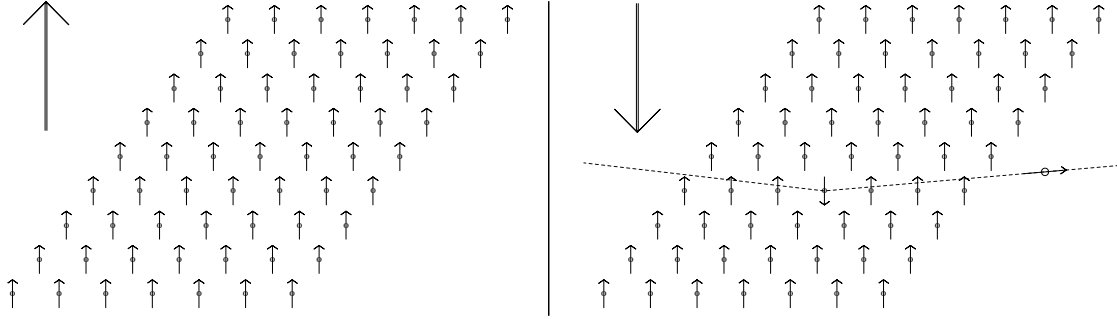


FIG. 30: Two-dimensional slice from an array of spins. On the left, before the apparatus is primed, the upward pointing field (indicated by the large arrow) orients the spins upward and the temperature is low enough so that opposite pointing spins are unlikely. On the right, the magnetic field has been reversed, so that all-spins-up is metastable, not stable. A particle has passed through the array, flipping one spin and being scattered slightly in the process. Subsequently the neighbors of the flipped spin will turn over because the reversed bond allows them to align with the field.

five opposed and now one neighbor favors the flip. Subsequent flips are yet more enhanced and the “droplet” grows.

As the Hamiltonian for the interactions affecting only the spins take

$$H_s = -J \sum_{\langle \ell \ell' \rangle} \sigma_{\ell z} \sigma_{\ell' z} + h \sum_{\ell} \sigma_{\ell z}, \quad (127)$$

where  $\ell$  and  $\ell'$  range over the spins and  $\langle \ell \ell' \rangle$  means summation over nearest neighbor pairs. (Take  $\hbar = 1$ .) For ferromagnetism, the spin-spin coupling constant  $J$  is positive. The external field is  $h$ . The operator  $\sigma_{\ell z}$  is the tensor product of the quantum spin operator  $\sigma_z$  for the  $\ell^{\text{th}}$  spin and the identity operator for the others.

Besides the spin degrees of freedom there is the underlying lattice whose phonons interact with the spins. Take the combined lattice and lattice-spin Hamiltonian to be

$$H_p = \sum_{k=1}^N \omega_k b_k^\dagger b_k + \sum_{\ell} \sum_{k=1}^N \left( \gamma_k e^{i\phi_k(\ell)} \sigma_{\ell+} b_k + \gamma_k^* e^{-i\phi_k(\ell)} \sigma_{\ell-} b_k^\dagger \right). \quad (128)$$

The operators  $b_k$  and  $b_k^\dagger$  are the phonon annihilation and creation operators for the  $k^{\text{th}}$  mode (out of  $N$ ) of the lattice. The mode's frequency is  $\omega_k$ . The phonons couple to the spins with coupling constants  $\gamma_k$ , together with a phase,  $e^{i\phi_k(\ell)}$ , that varies with the position of the spin on the lattice. The operator  $\sigma_{\ell+}$  is the raising operator for the  $\ell^{\text{th}}$  spin and it and  $\sigma_{\ell-}$  are defined analogously to  $\sigma_{\ell z}$ . (These interactions are completely standard and in a photon context give the Jaynes-Cummings Hamiltonian.)

Finally we present the portion of the Hamiltonian for the particle ( $\mathbf{x}$ ) to be detected. We assume that the effect of  $\mathbf{x}$  is to induce a strong coupling between the spin and the phonons. Its Hamiltonian is taken to be

$$H_d = \frac{p^2}{2m} + \sum_{\ell} v(x - x_{\ell}) \sum_{k=1}^N \left( c_k e^{i\phi_k(\ell)} \sigma_{\ell+} b_k + c_k^* e^{-i\phi_k(\ell)} \sigma_{\ell-} b_k^\dagger \right). \quad (129)$$

Again, the  $c_k$  are coupling constants and the function  $v(u)$  is non-zero only when  $|u|$  is near zero. We take  $v$  dimensionless and of order unity. The  $c_k$  are much larger than the usual lattice-spin coupling, i.e.,  $|c_k| \gg |\gamma_k|$ .

The total Hamiltonian for the system is the sum of the three pieces we have displayed,  $H_{\text{total}} = H_s + H_p + H_d$ . This model does not have an exact solution. But the point is not to provide exact solutions, only to arrive at the physical behavior, to show that notwithstanding an entirely quantum treatment, we have a system that behaves like a measuring apparatus. This is established in the following way. We first describe the response of the system qualitatively and then indicate what specific calculations are needed to justify the statements. Some of those calculations will be done here, some are in [1], some in [54]. Having done this, we will return to our primary goal, the exhibiting of special states.

### *Qualitative description of the operation of the detector*

Since  $v(\cdot)$  is short ranged, the impinging particle effectively interacts with only one spin. With this spin there is significant interaction for only a short time. As a result, once the particle is out of range of the spin, there is amplitude for having flipped the spin, amplitude for not having flipped. The wave function will have two components, not yet macroscopically distinct. In either case X moves away from the array and the next stage takes place by means of the coupling due to the  $\gamma_k$ . For the component that has not flipped, the system is more or less where it was before, and it is unlikely to flip. For the wave function component with the (single) flipped spin, the  $\gamma_k$ -coupling can either unflip the spin, or can cause neighboring spins to flip. Because the energy balance has shifted, the most likely thing to happen is the flipping of the neighbors of the spin that was hit by X. Once this occurs the tendency becomes overwhelming. Since unflipping was unlikely, even after a single spin flip, the probability of detection (by spin  $\neq 0$ ) will be the square of the amplitude for flipping after the single-spin encounter.

The most important calculations required for analysis of the model are 1) flip probability for a single scattering event; 2) false positives: probability of a flip in the absence of any external particle (i.e., via  $\gamma_k$  alone).

#### *1. Spontaneous spin flip: false positives and ground state stability*

False positives result from a decay process and are easy to calculate. Such spontaneous transitions can be computed using the formalism developed for decay in Sec. II. Because our Hamiltonian  $H_p$  only allows creation or annihilation of a single phonon, one can identify the matrix elements that correspond to the Hamiltonian given in Eq. (55). That form was

$$H = \begin{pmatrix} \nu & C^\dagger \\ C & \Omega \end{pmatrix}, \quad \psi = \begin{pmatrix} x \\ Y \end{pmatrix}. \quad (\text{Same as Eq. (55).}) \quad (130)$$

where the first component (“ $x$ ,” to be called component number “0”) is the initial state and the other components (“ $Y$ ”) are the decay products. We take as the initial state  $|+, n_1, n_2, \dots\rangle$ ; here the integers  $n_k$  are the eigenvalues of  $b_k^\dagger b_k$  and the “+” represents the state of the particular spin (from the array) on which we now focus. The diagonal matrix element of the Hamiltonian for this initial state has the value  $H_{00} = -6J + h + \sum_k n_k \omega_k$ . This state couples to all states of the form  $|-, n_1, n_2, \dots, (n_{k'} + 1), \dots\rangle$ ,  $k' = 1, 2, \dots$ . These

states have diagonal matrix elements  $H_{00} + 12J - 2h + \omega_{k'}$ . The coupling due to the terms  $\gamma_k e^{i\phi_k(\ell)} \sigma_{\ell+} b_k$  and their adjoints, give rise to  $\gamma_k e^{i\phi_k(\ell)} \sqrt{n_k + 1}$  in the first row of the Hamiltonian (to the right of the “00” diagonal term), and to its adjoint in the first column. In Sec. II we calculated the transition probability for the decay, which in this context is spin flip. The rate is given by the “Golden Rule,” as found in Eq. (67). To use that formula we note that the energy deposited in the phonon mode matches initial and final energies, namely,  $\Delta E \equiv 2h - 12J$ , which we here take to be slightly positive. Let  $\bar{k}$  be the phonon index for which  $\omega_{\bar{k}} = \Delta E$ , then in the continuum limit the “Golden Rule” gives

$$\Gamma = 2\pi \langle n_{\bar{k}} \rangle \rho(\bar{k}) |\gamma_{\bar{k}}|^2 \quad (131)$$

The factor  $\langle n_{\bar{k}} \rangle$  is the expected number of phonons in the initial state, and for inverse temperature  $\beta = 1/k_B T$  is given by  $[\exp(\beta \Delta E) - 1]^{-1}$ . For practical situations,  $\langle n_{\bar{k}} \rangle$  is of order unity. Under the assumed physical conditions, *the important factor in Eq. (131) is  $\rho(\bar{k})$ , the density of states.* For low energy phonons this is proportional to the square of the frequency; thus  $\Gamma \sim \Delta E^2 \ll 1$ . It follows that if the detector contains  $M$  spins it will be unlikely to have any spontaneous flips provided  $M \Delta E^2 |\gamma_k|^2 \ll 1$ . This means that our detector is not “macroscopic” in the sense of operating at *any* possible size, but is limited to dimensions that could be called mesoscopic. Once the signal is registered on our detector it can be recorded by a second level detector, just as the droplets in the cloud chamber are photographed before the next decompression/compression cycle.

## 2. Scattering-induced spin flips

For convenience we assume that  $\mathbf{x}$  traverses the detector rapidly, on a time scale such that little takes place due to the usual spin-phonon interaction (mediated by “ $\gamma_k$ ”) or among the phonons. Also we take the kinetic energy of  $\mathbf{x}$  to be large compared to the detector interaction; thus the bend in the trajectory in Fig. 30 is exaggerated. Finally, the range of the potential  $v$  is such that only a single spin is involved in the initial stage of detection. Let that spin be labeled “0” and suppose (without loss of generality) that  $\phi_k(0) = 0$  ( $\phi_k$  is the phase in Eq. (128)). Under these circumstances we can confine attention to this spin, for which the result of the interaction will be

$$\psi_0 \rightarrow \exp \left[ -i \int dt v(x(t) - x_0) \sum_{k=1}^N \left( c_k \sigma_{0+} b_k + c_k^* \sigma_{0-} b_k^\dagger \right) \right] \psi_0 \quad (132)$$

where  $\psi_0$  is the initial spin wave function of spin  $\neq 0$ , which is  $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$  (i.e., “up”). The function  $x(t)$  is the path of  $\mathbf{x}$ , which we assume unaffected by the interaction. We can now write  $T \equiv \int dt v(x(t) - x_0)$ , an effective contact time (since  $v$  is dimensionless). We are therefore interested in the unitary matrix

$$U = \exp \left[ -iT \sum_{k=1}^N \left( c_k \sigma_{0+} b_k + c_k^* \sigma_{0-} b_k^\dagger \right) \right] \quad (133)$$

We henceforth drop the subscript “0” referring to the spin number. To obtain an explicit form for  $U$ , define a new set of boson operators. Let

$$\beta_1 = \sum_k \frac{c_k}{\tilde{c}} b_k \quad \text{where } \tilde{c} \equiv \sqrt{\sum_k |c_k|^2} \quad (134)$$

(The “1” is the phonon-state label.) The other boson operators,  $\beta_2, \dots, \beta_N$ , are built from the  $b_k$  ( $k = 1, \dots, N$ ) that are independent of  $\beta_1$ . In terms of the  $\beta$ 's

$$U = \exp\left(-i\tilde{c}T(\beta_1\sigma_+ + \beta_1^\dagger\sigma_-)\right) \quad (135)$$

It is an exercise in spin matrices to evaluate  $U$ . With the definition

$$\hat{\nu}_1 = \beta_1^\dagger\beta_1 \quad (\text{an operator}) \quad (136)$$

we find [55]

$$\begin{aligned} U &\equiv \exp(-i\tilde{c}T(\beta_1\sigma_+ + \beta_1^\dagger\sigma_-)) \\ &= \sigma_+\sigma_- \cos\left(\tilde{c}T\sqrt{\hat{\nu}_1+1}\right) + \sigma_-\sigma_+ \cos\left(\tilde{c}T\sqrt{\hat{\nu}_1}\right) \\ &\quad -i \left\{ \frac{\sin\left(\tilde{c}T\sqrt{\hat{\nu}_1+1}\right)}{\sqrt{\hat{\nu}_1+1}}\beta_1\sigma_+ + \frac{\sin\left(\tilde{c}T\sqrt{\hat{\nu}_1}\right)}{\sqrt{\hat{\nu}_1}}\beta_1^\dagger\sigma_- \right\} \end{aligned} \quad (137)$$

Using  $U$  we immediately get the wave function at time  $t$ . Because of the spin index and the two different phonon bases that we use, there is a certain notational overload. From the form of  $U$  it is obvious that the eigenvalues of  $\hat{\nu}_k \equiv \beta_k^\dagger\beta_k$ , for  $k \neq 1$ , are of no importance for the action of  $U$ . Therefore in looking at the evolution of the spin-0 wave function and the phonon wave functions we specify only the spin state (of spin 0) and the number of phonons of type 1, i.e., the eigenvalue of  $\hat{\nu}_1$ . Let this eigenvalue be designated  $\nu_1$ . Thus if the initial spin-0-cum-phonon wave function is  $\psi(0) = |+, \nu_1\rangle$ , by Eq. (137) it evolves to

$$\begin{aligned} \psi(t) &= \cos\left(\tilde{c}T\sqrt{\nu_1+1}\right) |+, \nu_1\rangle - i \frac{\sin\left(\tilde{c}T\sqrt{\nu_1}\right)}{\sqrt{\nu_1}} \sqrt{\nu_1+1} |-, \nu_1+1\rangle \\ &= \cos\left(\tilde{c}T\sqrt{\nu_1+1}\right) |+, \nu_1\rangle - i \sin\left(\tilde{c}T\sqrt{\nu_1+1}\right) |-, \nu_1+1\rangle \end{aligned} \quad (138)$$

This is the state of spin 0 after  $\mathbf{x}$  has passed [56]. If the remaining processes are nearly certain (when spin 0 is down, the whole thing turns over, otherwise it does not) then Eq. (138) gives the detection and non-detection amplitudes. For a thermal distribution of phonons in the initial state, the probability of detection is

$$\text{Pr}(\text{detection}) = \frac{1}{\mathcal{Z}} \left\langle \sum_{\nu_1} \sin^2\left(\tilde{c}T\sqrt{\nu_1+1}\right) e^{-\beta\nu_1\Omega} \right\rangle_{\tilde{c}T}, \quad (139)$$

where  $\mathcal{Z} = 1/(1 - \exp(-\beta\Omega))$ ,  $\Omega = \sum_k |c_k|^2 \omega_k / \tilde{c}^2$ , and the subscript  $\tilde{c}T$  on the angular, averaging brackets indicates an average over particle-spin contact times ( $T$  is built from the trajectory of  $\mathbf{x}$ ) [57].

### 3. Operation of the detector

This is everything we need to know about the detector. After  $\mathbf{x}$  traverses the array, the wave function is of the form given in Eq. (138). The neighbors of spin 0 now flip due to their interaction with the phonons and driven by the external field. This flip rate is much greater than that for spontaneous flips. In three dimensions, flipping  $M$  spins takes on the



order of  $M^{1/3}$  flip times (at the fast, alignment-producing flip rate). If in this interval the probability of a spontaneous flip is small, then false positives are unimportant.

With  $M$  spins overturned, even for modest  $M$  (perhaps 1000 or fewer), we have entered the classical domain in two senses. 1) The overturning of the spins is irreversible. Poincaré recurrence times are enormous. 2) The disturbance is large enough so that subsequent evolution (e.g., a probe of the detector's magnetic moment) can be treated classically.

We next look at the wave function of the detector. After the first spin has completed its contact with  $\mathbf{X}$ , the detector wave function is given by Eq. (138) (or a superposition of such terms) crossed into the wave function of all the other spins and phonons. Continuing, the wave function of the entire system after a further time interval,  $\bar{t}$ , sufficiently long for a large number of spins to fall, is given by the action of  $\exp(-iH\bar{t})$  on the full detector state just described. Thus

$$\begin{aligned} \Psi = e^{-iH\bar{t}} & \left[ \cos(\tilde{c}T\sqrt{\nu_1+1}) |+, \nu_1\rangle - i \sin(\tilde{c}T\sqrt{\nu_1+1}) |-, \nu_1+1\rangle \right] \\ & \otimes \left| \text{spins other than } 0, \text{ all of which are up} \right\rangle \\ & \otimes \left| \text{phonons other than } \beta_1, \text{ all of which are unchanged} \right\rangle. \end{aligned} \quad (140)$$

As discussed above, the action of  $e^{-iH\bar{t}}$  leads to

$$\begin{aligned} \Psi = & \cos(\tilde{c}T\sqrt{\nu_1+1}) \left| \text{all spins up; phonons hardly changed} \right\rangle \\ & - i \sin(\tilde{c}T\sqrt{\nu_1+1}) \left| \text{spins down; phonons somewhat excited} \right\rangle \end{aligned} \quad (141)$$

This is a Schrödinger cat.

#### 4. Scattering-induced spin flips with SPECIAL initial conditions

Our explicit microscopic construction of a quantum measurement apparatus has had one purpose: to exhibit special states for that apparatus. After the passage of the particle  $\mathbf{X}$ , the wave function of spin  $\#0$  and the phonons was

$$\psi(t) = \cos(\tilde{c}T\sqrt{\nu_1+1}) |+, \nu_1\rangle - i \sin(\tilde{c}T\sqrt{\nu_1+1}) |-, \nu_1+1\rangle \quad (142)$$

where the only important fact about the phonon state is its quantum number with respect to a particular mode, “1”; the other modes can be doing anything, and are unaffected by  $\mathbf{X}$ . (With the departure of  $\mathbf{X}$  the energy is spread, since the mode “1” is not an eigenstate of the Hamiltonian.) Furthermore, we consider a situation where once this one spin is flipped the continuation of the process is inevitable, and had it not flipped, nothing would have happened.

Now  $T$  is a number that will vary widely, depending on the proximity of  $\mathbf{X}$  to the spin; there may also be variation in the coupling constants,  $c_k$ . In addition, from run to run there can be a wide range of values of  $\nu_1$ , representing thermal fluctuations in the lattice near the spin site 0. Suppose it should happen that  $\theta \equiv \tilde{c}T\sqrt{\nu_1+1} \approx \pi/2$ . Then (by Eq. (64)) the spin-up component of  $\psi(t)$  has coefficient  $\cos\theta$ , which is close to zero. The spin-down

component, by contrast, will have magnitude near 1. With these values of  $T$  and  $\nu_1$  the final apparatus state is *definite*. On the other hand, for another scattering event it may happen that  $\theta \approx \pi$ , so that despite the rather close coupling of scatterer and detector, there will be hardly any effect on the spin array. One spin does a handstand, but recovers completely. Again the final apparatus state is definite. As can immediately be seen, these are but two possibilities for *special* states, states for which the experiment gives a *definite* result and the system does *not* find itself in a Schrödinger cat state at the end of the experiment.

It follows that if it should happen that for a given passage of an X particle the initial state (of X and of the phonons) is an eigenstate of  $\hat{\nu}_1$  with an eigenvalue such that  $\theta (\equiv \tilde{c}T\sqrt{\nu_1 + 1})$  is close to  $[\text{integer}] \times \pi/2$ , then the final state will *not* be a superposition of two different macroscopic states, but only a single such state. This identifies a class of special states for the cloud chamber model apparatus.

### 5. Other apparatus models

In [1], several other apparatus models are studied and their special states shown to exist.

### C. Recovering the usual probabilities

How does probability enter classical physics? If you flip a coin, the probability of “heads” or “tails” is proportional to the region of phase space (of the complete system: coin, arm, air, anything else having influence) associated with each outcome. It will be the same in quantum mechanics. Since I have only deterministic (unitary) time evolution, this is the only possibility.

Given the vastness of phase space or Hilbert space (for macroscopic systems), once you have *any* special states, you should have many. Moreover, just as they are macroscopically indistinguishable from non-special states, so they are indistinguishable from each other. Two remarks before I state the probability rule: (1) By the correspondence principle, “volume of classical phase space region” becomes “dimension of Hilbert subspace;” (2) if two special states have the same macroscopic outcome, so does any linear combination of them, so each outcome can be associated with a Hilbert subspace.

This then is the rule for quantum probabilities: *the probability of a particular outcome is proportional to the dimension of the subspace of Hilbert space associated with that outcome.*

This is a far-reaching postulate. Given the difficulty of exhibiting even one class of special states, it would seem impossible to classify all special apparatus states and on top of that guarantee that their abundance exactly matches the absolute value squared of the amplitudes of the system being measured.

Let me mention my experience in confronting this issue. At first I didn’t. Then at some point I was explaining these ideas, and persistent questioning [58] made me realize that without cleaning this up I was indulging in daydreams. I will explain my general reasoning below, but I should say that after a week or so of thinking about this I had reached an impasse, a contradiction, and felt that my all my work on the quantum measurement problem had indeed collapsed into daydreams. But there was a way, and when I found it, it had a certain cleanness to it (perhaps I should leave that to others to judge).

The goal then is to take a general approach, one that should not depend too closely on the apparatus. For specificity though, imagine a Stern-Gerlach measurement of a spin-1/2

particle. Suppose it is in a superposition of the following sort:

$$u_\theta = e^{i\theta\sigma_x/2} \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \quad (143)$$

This is the state in which it is prepared. The record of its arrival at a screen downstream from the magnets will show an event in one of two regions, one for spin-up, the other for spin-down. There could be unusual microstates *anywhere* along the way (that is easy to understand if the rationale for special states is future conditioning). The screen could strangely not function correctly, the magnetic field could experience a fluctuation, or the single spin-bearing particle could be subject to stray fields or other influences at any time after its preparation. As usual when looking for something unusual (the special state in this case) one should look for the *least* unusual. The measure may be small, but it would be bigger than the alternatives. It seems to me that acting unusually on a single spin is easier, i.e., less unlikely, than acting on all the electrons that give rise to the field in the big magnet, or persuading the many atoms involved in detection in the screen to fire or not to fire.

Before giving the detailed argument I remind you of the expected result. After passing through the magnets, the amplitude for arriving at the UP detector is proportional to  $\cos(\theta/2)$ , and for arriving at the DOWN detectors is proportional to  $\sin(\theta/2)$ . The ratio of DOWN to UP detection is therefore  $\tan^2(\theta/2)$ .

Since we are looking for a general result, we take the perspective of the “system,” the particle bearing the spin. Whatever the features of the special state it experiences, they will look like perturbations, disturbances, as it makes its way from preparation onward. If those disturbances could rotate it from the angle  $\theta$  in Eq. (143), to the angle 0 or  $\pi$ , before it reached the magnet, the special state would have done its job. So without going into the details of the cause of these disturbances we imagine that there is some likelihood of one or another disturbance. Let us call such a disturbance a “kick” and suppose that there is some kick distribution function,  $f(\psi)$ , with  $f$  the relative probability for a kick that will change the angle  $\theta$  in Eq. (143) to  $\theta - \psi$  [59]. (It is worth elaborating on the relation of this to the exact “apparatus” [60] microstates that I earlier called  $\Omega$ . The idea is that there are many  $\Omega$ ’s, some of which do not kick the particle spin (so for them  $\psi = 0$ ), some of which give a big kick, etc. The probability that I associate with a given  $\psi$  is the relative number of Hilbert space dimensions in the overall space of such  $\Omega$ ’s that move  $\theta$  by the amount  $\psi$ . Thus  $\Omega$  provides a fully quantum description of the environment, de facto to be considered part of what we call apparatus.)

I have now gone far enough into the story to explain why I thought my theory was washed up. Clearly,  $f(\psi)$  should be largest at  $\psi = 0$ , and drop off for larger  $|\psi|$ . So one would expect to use a succession of small kicks to get from  $u_\theta$  to  $u_0$  or  $u_\pi$ . By the central limit theorem, this would require that  $f$  be a Gaussian. But you cannot get the ratio  $\tan^2(\theta/2)$  from a Gaussian. At some point it dawned on me that maybe the assumptions of the central limit theorem should be dropped, mainly the existence of a second moment for the distribution. So I’ll turn the question around and ask, what function  $f$  can give me the ratio I need, and then worry about its moments.

To get to the UP state you need a kick by  $\theta$  or by  $\theta + 2\pi$  or  $\theta + 4\pi$ , etc. Similarly for the DOWN state you need a kick by  $\pi + \theta$ , or that, plus multiples of  $2\pi$ . It is useful to define

$$F(\psi) = \sum_{k=-\infty}^{\infty} f(\psi + 2k\pi). \quad (144)$$

The demand that the ratio of probabilities recover the standard value is the functional equation

$$\tan^2 \frac{\theta}{2} = \frac{F(\theta + \pi)}{F(\theta)}. \quad (145)$$

To guess a solution of this equation consider  $\theta$  near zero. Presumably in this case  $F$  should be dominated by the  $k = 0$  term and the numerator in Eq. (145) should not be doing anything interesting. Therefore one expects  $\theta^2 \sim 1/f(\theta)$ . It turns out this is an exact solution! The following is an identity:

$$\tan^2 \frac{\theta}{2} = \frac{F(\theta + \pi)}{F(\theta)} \quad \text{with} \quad F(\theta) = \sum_{n=-\infty}^{\infty} \frac{1}{(\theta + 2n\pi)^2} \quad (146)$$

The physics therefore requires that  $f = \text{const}/\theta^2$ , down to  $\theta$  as small as has been experimentally checked. For finiteness however this cannot hold all the way to zero. To find a good form for the cutoff, I review a bit of mathematics.

As a probability distribution (and irrespective of its small  $\theta$  behavior)  $f$  does not have a second moment, and is just borderline for a first moment. The study of distributions that do not satisfy the central limit theorem was pioneered by Paul Lévy. Just as the Gaussian is a kind of basin of attraction (under summation) for random variables having second moments, so the Lévy distributions are attractors for those lacking such moments. The canonical form for distributions with  $f \sim 1/x^2$  is the Cauchy distribution (a particular Lévy distribution)

$$C_a(x) = \frac{a/\pi}{x^2 + a^2}. \quad (147)$$

We therefore take  $f$  to be the Cauchy distribution with small parameter  $a$ .

These distributions have lovely and surprising properties. For our purposes I emphasize the tendency to have large fluctuations. For example, it is about 5 times less likely to get the result  $10a$  than the result  $a$  from the Cauchy distribution. For a Gaussian with standard deviation  $a$  it is about  $3 \times 10^{21}$  times less likely. For other comparisons see [1].

I will return to the parameter  $a$  when discussing experimental tests (Sec. XD).

The story that I have just told for 2-state systems goes through for any finite number of components. The “kicking” takes place in a larger Hilbert space, but the mathematics, including the role of the Cauchy distribution, goes through remarkably smoothly.

### *|\psi|^2 as probability*

A problem of great interest in quantum mechanics is whether it is possible to derive the association of  $|\psi|^2$  and probability. For example there was heated discussion of this issue at a recent Séminaire Poincaré [61].

The derivation just given, in which the Cauchy distribution provides just the right asymptotics to recover the  $|\psi|^2$  relation, does *not* go through for any other power of  $|\psi|$ . This is not a proof that in the Copenhagen approach you must get  $|\psi|^2$ . Rather it’s the statement that if you could derive the kicks from an underlying fluctuation or large deviation theory of matter, you would also be deriving the usual relation of probability and wave function.

### D. Experimental tests

There are two categories of test: (1) confounding Nature’s efforts to provide a special state, and (2) noticing the Cauchy noise.

It would seem easy to isolate an experiment so as to prevent outside influences from sending the microstate into one of the possible macroscopic outcomes (in which case, a measurement of the system in that outcome state would disprove my theory). But in fact it is *not* easy. Even with very few atoms around, if they act coherently their impact can be great. In [1] I estimate the level of vacuum needed in a typical scattering experiment and find it to be extreme. EPR experiments seem more promising, since one can form a better idea of where the “specializing” must take place. In any case, I have not yet formulated a sharp test of this sort. I have also begun to worry lately that dark matter may have something to do with this, in which case isolation would be impossible.

The Cauchy noise however might nevertheless reveal itself. This could happen by studying fluctuations in ordinary matter. After all, the functional form of the Cauchy distribution is well-known in physics: it’s just the Lorentz line shape. So there are many physical variables distributed in this way.

A second manifestation of the noise comes from the parameter “ $a$ ” in the Cauchy distribution,

$$C_a(x) = \frac{a/\pi}{x^2 + a^2}. \quad (148)$$

To achieve perfect matching with what seem to be physical demands, one needed to satisfy Eq. (145). Taking  $f(\theta) = C_\gamma(\theta)$ , such agreement would require  $\gamma = 0$ . It’s easy to calculate the probabilities if  $\gamma \neq 0$ . We need the sum

$$\begin{aligned} F(\theta) &= \frac{\gamma}{\pi} \sum_{n=-\infty}^{\infty} \frac{1}{(\theta + 2n\pi)^2 + \gamma^2} \\ &= \frac{1}{2\pi} \operatorname{Im} \left( \frac{1}{\tan \frac{1}{2}(\theta - i\gamma)} \right) = \frac{(\frac{1}{2\pi}) \tanh \frac{1}{2}\gamma}{\sin^2 \frac{1}{2}\theta + \cos^2 \frac{1}{2}\theta \tanh^2 \frac{1}{2}\gamma} \end{aligned} \quad (149)$$

For an initial state  $u_\theta$ , the probability of observing UP is

$$\operatorname{Pr}(\text{UP}) = \frac{F(\theta)}{F(\theta) + F(\theta + \pi)} = \frac{\cos^2 \frac{\theta}{2} + \sin^2 \frac{\theta}{2} \tanh^2 \frac{\gamma}{2}}{1 + \tanh^2 \frac{\gamma}{2}} \quad (150)$$

For small  $\gamma$  this gives

$$\operatorname{Pr}(\text{UP}) \cong \cos^2 \frac{\theta}{2} - \frac{\gamma^2}{4} \cos \theta \quad (151)$$

This is *not* the standard result. An atom going in with spin up ( $\theta = 0$ ) has probability  $\gamma^2/4$  of being misread as down. If such an effect were detected, one could further get some idea (from the size of  $\gamma$ ) of the physical origins of this noise.

**Remark:** There is already a source of small-angle error that is known from the “WAY” theorem [62–64]. Because the quantity measured is conserved by the total Hamiltonian, the work of Wigner, Araki and Yanase implies that there is necessarily a small error, on the order of the ratio of the angular momentum to be measured to that of the apparatus. Nevertheless, by selecting a suitable experimental context, the search for small angle deviations seems a promising avenue.

**E. Excerpt from Ref. [1] (*Time's Arrows and Quantum Measurement*), Sec. 6.3, on the determinism implicit in this worldview.**

*This excerpt emphasizes the fact that the quantum measurement theory described in these notes (and in the book) require a deterministic world. The entire history of the universe, past and future, is fixed.*

The other claim is the more dramatic. First you can't control your initial state, and second it is always one of the presumably rare 'special' ones. For those who have read the earlier chapters of this book, this claim should not be surprising. The thermodynamic arrow of time is what gives us our strong prejudice on the arbitrariness of initial conditions and the specificity of final conditions. In a dynamical system for which the boundary conditions are naturally defined at more than one time, the selection of allowable states (from among all candidates macroscopically possible) will not favor initial conditions in this way.

Therefore we can already anticipate that the constraint on initial states will have something to do with the giving of conditions in the future, conflicting with our primitive intuition that only statements about the past influence the present state of a system.

There is another claim implicit in this proposal. Suppose your laboratory is wealthy enough to have two cloud chambers and before doing the experiment you have to decide between them, based perhaps on choosing the one with a gas leak or the one with an unreliable compressor. (Well, maybe the lab is not so wealthy after all!) The chamber that you opt to use will turn out to be the one with the right 'special' state. Or suppose you decide to aim the beam a little differently. Then it will be other gas molecules that are primed for perfect detection or perfect non-detection. It follows that the 'special' states that occur are coordinated with your decision. But since (according to this theory) nothing ever happens except pure, unitary quantum evolution, the precursors of these states were heading where they were going before you made your 'decision.' So your decision was not a decision, and your wave function and that of the detectors are correlated. Pursuing this line of reasoning to ever greater scales, it follows that my ideas can only be valid if there is a single wave function for the entire universe. This wave function has the precise correlations necessary to guarantee 'special' states at every juncture where they are needed.

Again, it is my hope that the edge on the foregoing assertion has been taken off by the earlier parts of this book in which I discussed the arrow of time and past and future boundary conditions. A future boundary condition can trivially generate long range correlations. For chaotic systems these correlations demand extreme precision. What the foregoing discussion implies is that if I propose to motivate the appearance of 'special' states by a future boundary condition, that boundary condition should involve the entire universe. So if I am right about this explanation of the quantum measurement problem, not only must we reexamine statistical mechanics, but cosmology plays a role as well.

*See also the notes to Sec. 6.3 of the book (pp. 220–221).*

### Acknowledgments

I thank the participants in this course for many helpful comments. The support of the NSF (most recently PHY 05 55313) is gratefully acknowledged.

## References and endnotes

- [1] L. S. Schulman, *Time's Arrows and Quantum Measurement* (Cambridge University Press, New York, 1997), see <http://www.cambridge.org/us/catalogue/catalogue.asp?isbn=0521567750>.
- [2] K. Huang, *Statistical Mechanics* (Wiley, New York, 1963).
- [3] R. E. Ulanowicz, *Quantitative methods for ecological network analysis*, *Comp. Bio. Chem.* **28**, 321-339 (2004).
- [4] M. Kac, *Some Stochastic Problems in Physics and Mathematics*, Colloquium Lectures in the Pure and Applied Sciences, given at the Field Research Laboratory, Socony Mobil Oil Co., Inc., 1956.
- [5] M. Kac, *Probability and Related Topics in Physical Sciences* (Interscience Publishers, Ltd, London, 1959), proceedings of the Summer Seminar, Boulder, Colorado, 1957. With special lectures by G. E. Uhlenbeck, A. R. Hibbs and B. van der Pol.
- [6] L. Boltzmann, *Lectures on Gas Theory* (Dover Publications, New York, 1995), trans. by S. G. Brush. See §90.
- [7] A. Katz, *Principles of Statistical Mechanics: The Information Theory Approach* (Freeman, San Francisco, 1967).
- [8] L. S. Schulman, *Note on the Quantum Recurrence Theorem*, *Phys. Rev. A* **18**, 2379-2380 (1978).
- [9] The eigenvectors given in [1] are incorrect.
- [10] L. S. Schulman, M. Stone, and D. J. Wallace, *Metastable Lifetimes, Level Crossings and Vertical Cuts in Quantum Mechanics*, *J. Phys. A* **11**, 1933-1941 (1978).
- [11] L. S. Schulman, *Unstable Particles and the Poincaré Semigroup*, *Ann. Phys.* **59**, 201-218 (1970).
- [12] A. Bohm and Y. Sato, *Relativistic Resonances: Their masses, widths, lifetimes, superposition, and causal evolution*, *Phys. Rev. D* **71**, 085018 (2005).
- [13] A. Bohm, M. Loewe, S. Maxson, P. Patuleanu, C. Puntmann, and M. Gadella, *Gamow-Jordan vectors and non-reducible density operators from higher-order S-matrix poles*, *J. Math. Phys.* **38**, 6072 (1997).
- [14] A. R. Bohm, H. Kaldass, and S. Wickramasekara, *Time Asymmetric Quantum Theory—III Decaying states and the Poincaré semigroup*, *Fortschr. Physik* **51**, 604-634 (2003).
- [15] The sign in the exponent in Eq. (2.5.15) of [1] is wrong, so that what you actually need is  $\text{Im}\omega > 0$ .
- [16] P. Facchi, H. Nakazato, and S. Pascazio, *From the Quantum Zeno to the Inverse Quantum Zeno Effect*, *Phys. Rev. Lett.* **86**, 2699-2703 (2001).
- [17] M. C. Fischer, B. Gutiérrez-Medina, and M. G. Raizen, *Observation of the Quantum Zeno and Anti-Zeno effects in an unstable system*, *Phys. Rev. Lett.* **87**, 040402 (2001).
- [18] L. S. Schulman, *Continuous and pulsed observations in the quantum Zeno effect*, *Phys. Rev. A* **57**, 1509-1515 (1998).
- [19] R. G. Newton, *Scattering Theory of Waves and Particles* (Springer, New York, 1982), 2nd ed.
- [20] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley, New York, 1991).
- [21] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge Univ. Press, Cambridge, UK, 2000).
- [22] B. Gaveau and L. S. Schulman, *Master equation based formulation of non-equilibrium statistical mechanics*, *J. Math. Phys.* **37**, 3897-3932 (1996).
- [23] The non-uniqueness in retrodicting is related to the statistician's ambiguities in constructing

- estimator functions. See [65].
- [24] L. S. Schulman, *A Computer's Arrow of Time*, *Entropy* **7**, 221-233 (2005), originally presented at the conference, "The direction of time," ZiF, Bielefeld, January 2002, organized by P. Blanchard and S. Albeverio.
- [25] R. Carter, *Exploring Consciousness* (Univ. California Press, Berkeley, 2002).
- [26] E. Schrödinger, *What is Life?* (Cambridge Univ. Press, Cambridge, 1992), Canto reprint ed.
- [27] T. Gold, *The Arrow of Time*, in *Proceedings of the 11th Solvay Conference, "Structure and Evolution of the Universe"* (Stoops, Brussels, 1958), see also [66, 67].
- [28] L. S. Schulman, *Causality is an effect*, in *Time's Arrows, Quantum Measurements and Superluminal Behavior*, edited by D. Mugnai, A. Ranfagni, and L. S. Schulman (Consiglio Nazionale delle Ricerche (CNR), Rome, 2001), pp. 99–112, on the arXiv at cond-mat/0011507.
- [29] J. A. Wheeler, *Frontiers of time*, in *Problems in the Foundations of Physics*, edited by G. T. di Francia (North Holland, Amsterdam, 1979).
- [30] J. A. Wheeler, *Conference Summary: More Results than Ever in Gravitation Physics and Relativity*, in *General Relativity and Gravitation*, edited by G. Shaviv and J. Rosen (Wiley, New York, 1975), proceedings of the Seventh International Conf. (GR7). Tel-Aviv Univ., 1974.
- [31] L. S. Schulman, *Illustration of Reversed Causality with Remarks on Experiment*, *J. Stat. Phys.* **16**, 217-231 (1977).
- [32] J. A. Wheeler, *Time Today*, in *Physical Origins of Time Asymmetry*, edited by J. J. Halliwell, J. Perez-Mercader, and W. H. Zurek (Cambridge Univ. Press, Cambridge, 1994).
- [33] L. S. Schulman, *Time Reversal for Unstable Particles*, *Ann. Phys.* **72**, 489-496 (1972).
- [34] I suppress the implicit time in the notation because there are too many "t"'s around.
- [35] R. Baierlein, *Thermal Physics* (Cambridge Univ. Press, Cambridge, 1999).
- [36] B. Gaveau and L. S. Schulman, *Dynamical distance: coarse grains, pattern recognition, and network analysis*, *Bull. Sci. math.* **129**, 631-642 (2005).
- [37] L. D. Landau and E. M. Lifshitz, *Statistical Physics*, vol. 5 of *5th* (Pergamon Press, Oxford, 1980), Course of Theoretical Physics.
- [38] F. R. Gantmacher, *The Theory of Matrices*, vol. I (Chelsea, New York, 1959).
- [39] F. R. Gantmacher, *Theory of Matrices*, vol. II (Chelsea, New York, 1959).
- [40] C. J. Thompson, *Mathematical Statistical Mechanics* (Princeton Univ. Press, Princeton, 1979).
- [41] B. Gaveau and L. S. Schulman, *Theory of non-equilibrium first order phase transitions for stochastic dynamics*, *J. Math. Phys.* **39**, 1517-1533 (1998).
- [42] B. Gaveau and L. S. Schulman, *Multiple phases in stochastic dynamics: geometry and probabilities*, *Phys. Rev. E* **73**, 036124 (2006).
- [43] J. Hawkins, *On Intelligence* (Times Books, Henry Holt & Co., New York, 2004), with S. Blakeslee.
- [44] W. W. Zachary, *Information-flow model for conflict and fission in small-groups*, *J. Anthropol. Res.* **33**, 452-473 (1977).
- [45] J. J. Halliwell, *Decoherent Histories and the Emergent Classicality of Local Densities*, *Phys. Rev. Lett.* **83**, 2481-2485 (1999).
- [46] J. J. Halliwell, *The Emergence of Hydrodynamic Equations from Quantum Theory: A Decoherent Histories Analysis*, *Found. Phys. Lett.* **39**, 1767-1777 (2000).
- [47] J. M. Rost and J. S. Briggs, *On the derivation of the time-dependent equation of Schrödinger*, *Found. Phys.* **31**, 693-712 (2001).
- [48] A. Allahverdyan, R. Balian, and T. M. Nieuwenhuizen, *Curie-Weiss model of the quantum measurement process*, *Europhys. Lett.* **61**, 452-458 (2003).



- [49] A. E. Allahverdyan, R. Balian, and T. M. Nieuwenhuizen, *Determining a Quantum State by Means of a Single Apparatus*, Phys. Rev. Lett. **92**, 120402 (2004).
- [50] K. Gottfried, *Quantum Mechanics* (Benjamin, New York, 1966).
- [51] B. Rossi, *High-Energy Particles* (Prentice-Hall, New York, 1952).
- [52] Recall the concept of *critical droplet*. Although the liquid has a lower free energy, the advantage of being in the liquid state is a per-unit-volume (bulk) effect. For a volume or droplet that is too small, the gain does not compensate for the cost in surface energy that goes with creating the droplet. The droplet size for which these effects balance is called “critical.” To condense to the liquid state you must form such a droplet. Only then will there be a tendency for the droplet to grow.
- [53] I don’t care whether or not there is a “true” phase transition in this system. All I need is a metastable, long-lived state with a long enough correlation length among the spins to freeze them in place for the required time interval. For metastable systems, the infinite volume “thermodynamic limit” is not a suitable concept.
- [54] B. Gaveau and L. S. Schulman, *Model apparatus for quantum measurements*, J. Stat. Phys. **58**, 1209-1230 (1990).
- [55] In Eq. (137)  $(\sin x)/x$  stands for an operator power series.
- [56] Recall that this passage is rapid on the time scale of the phonon modes—which is why Eq. (133) well approximates the propagator during the passage.
- [57] In [1], Eq. 6.2.13 is incorrect and should be replaced by Eq. (139).
- [58] My interrogator (if I recall correctly) was Jan Adrian Leegwater, then a student in Utrecht.
- [59] More precisely,  $f$  is the distribution function for the combined effect of all kicks sustained by the system on its way to the magnet.
- [60] “Apparatus” means equipment, environment, whatever can affect the particle.
- [61] B. Duplantier, J.-M. Raimond, and V. Rivasseau, *Poincaré Séminaire 2005* (Birkhäuser, Basel, 2006), articles by W. H. Zurek, J.-M. Raimond, S. Haroche, J. Kempe, D. Esteve, H. D. Zeh and E. Joos.
- [62] H. Araki and M. M. Yanase, *Measurement of Quantum Mechanical Operators*, Phys. Rev. **120**, 622-626 (1960).
- [63] E. P. Wigner, *Die Messung quantenmechanischer Operatoren*, Z. Physik. **133**, 101 (1952).
- [64] M. M. Yanase, *Optimal Measuring Apparatus*, Phys. Rev. **123**, 666-668 (1961).
- [65] L. S. Schulman, R. G. Newton, and R. Shtokhamer, *Model of Implication in Statistical Mechanics*, Phil. Sci. **42**, 503-511 (1975).
- [66] T. Gold, *The Arrow of Time*, Amer. J. Phys. **30**, 403-410 (1962).
- [67] T. Gold, *The Nature of Time* (Cornell Univ. Press, Ithaca, New York, 1967).