SECOND LECTURE

Now I want to find the average:

$$\left\langle \frac{1}{n}\left[N_b(t)-N_w(t)\right]\right\rangle_{\{S\}} = \left\langle \frac{1}{n}\sum_p \varepsilon_{p-1}\varepsilon_{p-2}\cdots\varepsilon_{p-t}\right\rangle_{\{S\}} \tag{17}$$

Whatever else about averages can be said, an average of a sum is always the sum of the averages. So now I put the average inside the sum (it will be understood that we are averaging over all positions of S, so I can suppress the symbol S.):

$$\left\langle \frac{1}{n}\left[N_b(t)-N_w(t)\right]\right\rangle = \frac{1}{n}\sum_p \left\langle \varepsilon_{p-1}\varepsilon_{p-2}\cdots\varepsilon_{p-t}\right\rangle \tag{18}$$

All you have to notice now is that all these averages are the same. Because it doesn't really matter what p is. You can simply take any point -- and moreover you can run the summation in the opposite order. Therefore,

$$\left\langle \frac{1}{n}\left[N_b(t)-N_w(t)\right]\right\rangle = \left\langle \varepsilon_1\varepsilon_2\cdots\varepsilon_t\right\rangle \tag{19}$$

How to calculate this average? So far, I haven't done anything -- I simply used the most rudimentary properties of the averaging operation. But now we must define a little bit of what we mean by this. We mean by this the following: That all the ε_p are either +1 or -1 subject to a condition. They are subject to a condition because I know that I have exactly m elements in this set S. So what is the condition? It is that if I sum all the ε_p on p then I get m minuses and n-m pluses, or a total of n-2m. Thus the condition which expresses the fact that I have exactly m elements in the set S is

$$\sum_p \varepsilon_p = n-2m \tag{20}$$

We can now say what the average means. It simply means to sum over all possible

sequences $\varepsilon_1, \varepsilon_2, \cdots \varepsilon_t$ subject to condition (20) and then divide by the number of all possible sequences. That's the definition of the average.

The number of possible sequences is very easy to get because I have m elements to choose out of n objects. That's the binomial coefficient $\binom{n}{m}$, so we have the denominator out of the way. The numerator is a little bit more complicated. It is the quantity $\varepsilon_1 \varepsilon_2 \cdots \varepsilon_t$ summed over all possible sequences subject to the condition (20). I will now show you a trick for evaluating this sum. Notice that it is the condition on the ε_p which is the nuisance. Because if we were free to sum over all possible sequences, then it would be very, very easy. In fact, it would be extraordinarily easy because you would simply have ε_1 summed over all possible values, times ε_2 summed over all possible values, times ε_3 etc. Since the ε_p can only be either plus or minus one, these sums are trivial.

Our trouble then is with this condition (20). This type of thing occurs over and over again in all probability problems -- this is a very elementary example of it. We use the following trick: Noticing that all numbers involved here are integers, consider the following formula:

$$\frac{1}{2\pi i} \oint \frac{dz}{z^{l+1}} = \begin{cases} 0 & \text{if } l \neq 0 \\ 1 & \text{if } l = 0 \end{cases} \tag{21}$$

This is simply a standard formula in complex variable theory written in a somewhat un-standard form. The contour of integration is a circle around the origin. Now in this formula, put $l = \sum_p \varepsilon_p - n + 2m$. Let us then write down this formula:

$$\frac{1}{2\pi i}\oint \frac{dz}{z^{\varepsilon_1+\varepsilon_2+\cdots+\varepsilon_n-n+2m+1}} = \begin{cases} 0 \text{ if } \sum_p \varepsilon_p \neq n-2m \\ 1 \text{ if } \sum_p \varepsilon_p = n-2m \end{cases} \tag{22}$$

But now the sum I am seeking is just

$$\sum_{\text{over all } \varepsilon's} \varepsilon_1\varepsilon_2\cdots\varepsilon_t \frac{1}{2\pi i}\oint \frac{dz}{z^{\varepsilon_1+\cdots+\varepsilon_n-n+2m+1}} \tag{23}$$

Where now the sum is over all epsilons, since our nuisance condition is now taken care of by the integral.

And now you interchange the integral with the summation:

$$\frac{1}{2\pi i}\oint \frac{dz}{z^{2m-n+1}} \sum_{\text{over all } \varepsilon's} \frac{\varepsilon_1\varepsilon_2\cdots\varepsilon_t}{z^{\varepsilon_1+\cdots+\varepsilon_n}} \tag{24}$$

And the nice thing is that this sum I can calculate very easily. Because you see that the general term in this summation can be written as follows:

$$\left(\prod_{k=1}^{t} \varepsilon_k / z^{\varepsilon_k}\right)\left(\prod_{k=t+1}^{n} 1/z^{\varepsilon_k}\right) \tag{25}$$

Where I assume that t is less than n. The general term splits up into a product of factors -- and the factors all depend on different subscripts. So you can simply sum each factor separately, and then multiply them out. That is, you can first perform the summation over ε_1 , then over ε_2 , etc. Now what happens if you sum the first factor over ε_1 ? ε_1 is either plus one or minus one, so the result is simply $(\frac{1}{z} - z)$. How many such things are there? All together, t of them. The last factor will behave differently,

because there is no ε in the numerator. It gives $(\frac{1}{z} + z)$, and there are n-t such terms. So all the sum simply becomes $(\frac{1}{z}-z)^t (\frac{1}{z}+z)^{n-t}$. And now I can express my sum as a very neat integral:

$$\frac{1}{2\pi i}\oint \frac{dz}{z^{2m-n+1}}\left(\frac{1}{z}-z\right)^t\left(\frac{1}{z}+z\right)^{n-t} \tag{26}$$

This is a very simple, but a very common trick, and it is important to understand all the steps. Now when you clean this up a little bit algebraically you get

$$\frac{1}{2\pi i}\oint \frac{1}{z}\left(\frac{1-z^2}{1+z^2}\right)^t \frac{(1+z^2)^n}{z^{2m}}\,dz \tag{27}$$

Remember that I must take this expression and divide the whole thing by $\binom{n}{m}$. Now as a matter of fact, we can also write a very complicated formula for $\binom{n}{m}$ which will be of the same form. Indeed,

$$\binom{n}{m} = \sum_{\text{over all } \varepsilon_s} \frac{1}{2\pi i}\oint \frac{dz}{z^{\varepsilon_1+\cdots+\varepsilon_n-n+2m+1}} \tag{28}$$

Because I'm simply counting each time $\sum_p \varepsilon_p$ is n-2m, and otherwise I don't count at all. The expression on the right is the same as formula (23) when t = 0. So it ought to be that

$$\binom{n}{m} = \frac{1}{2\pi i}\oint \frac{1}{z}\frac{(1+z^2)^n}{z^{2m}}\,dz \tag{29}$$

which is simply what you get from formula (27) by putting t = 0. And it is

very easy to verify that this is indeed so. I will now write out, in its full glory, the average which we want:

$$\left\langle \frac{N_b(t)-N_w(t)}{n} \right\rangle = \frac{\oint \frac{1}{z}\left(\frac{1-z^2}{1+z^2}\right)^t \frac{(1+z^2)^n}{z^{2m}}\,dz}{\oint \frac{1}{z}\,\frac{(1+z^2)^n}{z^{2m}}\,dz} \tag{30}$$

That is a nice compact way of writing it, if you like it. But it also gives me a way to calculate. It will be a nice simple exercise in the method of steepest descent. I am interested in the case where n is very large, and I will assume that $\frac{m}{n}$ approaches μ as n approaches infinity. Moreover, I assume that 2μ is less than one. Now I hope you are familiar with the principle of the method of steepest descent. We will apply it first to the integral in the numerator. The first step is to write the integrand in exponential form. It will be enough to consider only the factor that involves n and m. The rest is a perfectly well defined function which doesn't change with n and m. So we have that

$$\frac{(1+z^2)^n}{z^{2m}} = e^{n\log(1+z^2)-2m\log z} \tag{31}$$

There is no point in worrying about the branch of the logarithm -- it doesn't affect the result. Now we must find the place where the derivative of the exponent is zero -- that is the saddle point. Now what is the derivative of this? It is $\frac{2nz}{1+z^2} - \frac{2m}{z}$. For this to be zero, simply means that $\frac{z_0^2}{1-z_0^2} = \frac{m}{n}$, which I can say is μ . (I may as well assume it is exactly equal to μ). Now if I subtract from one, I get $1-\mu = \frac{1}{1+z_0^2}$ -- and I think that is all I

need to find Z_0. There are two values and we will take the positive one, although it doesn't matter. Now, having found the saddle point, you substitute for Z wherever you see it, the value of the saddle point.

What I need to calculate now is the saddle point for the integral in the denominator. But the equation determining the saddle point is just the one we have already written. The part of the integrand involving n and m has the same asymptotic behavior as before. The $\frac{1}{Z}$ is also exactly the same. So all that is left is simply the value of $\left[(1-Z^2)/(1+Z^2)\right]^t$ at the saddle point. And what is it? It is just $(1-2\mu)^t$. Hence, by the saddle point method the asymptotic behavior is

$$\left\langle \frac{N_p(t)-N_w(t)}{n} \right\rangle \sim (1-2\mu)^t \qquad (32)$$

This simply means that the average over all possible positions of the set S -- if these positions are assumed equally probable -- agrees exactly with what we had before. However, notice that in performing this limiting operation, I make certain assumptions. I kept t fixed while n and m went to infinity. The important thing is that t was fixed while n went to infinity, which in terms of physics means that t must be small compared to n. Otherwise you could not use the method of steepest descent. If t were of the order of magnitude of n, then of course, the whole thing would not be justified. I would have to include the factor involving t in writing out the exponent, and the position of the saddle point would be entirely different -- the asymptotic behavior would be entirely different.

So it's only if t is small compared to n -- or better yet, if t is fixed while the number of points tends to ∞ our statement can be maintained.

Now this is completely in agreement with what is usually thought in statistical mechanics and kinetic theory, that you can only believe in the conclusions of kinetic theory or statistical mechanics if the times of observation are short compared to the underlying Poincare' cycle. (Remember I spoke earlier of Poincaré's theorem; the Poincare' theorem tells me that I eventually have to come back close to my starting point in phase space. The mean time I have to wait before coming back -- this can be very long -- is called Poincare' cycle). In our model, the Poincare' cycle is 2n because the whole thing is periodic with this period.

Let us think a little bit about what all this means. Suppose I'm going to plot $\frac{1}{n}\left[N_b(t)-N_w(t)\right]$ against t. How many curves am I going to get? I am going to get $\binom{n}{m}$ curves because I get a different one for each choice of the set S. Each of these curves starts always with one, and is periodic with period 2n. Now I am going to fix myself a t -- I'm going to look at these curves only at one point. Think of n as being very large compared to t -- think of n being 10^{23}, and of t being 10^6. Now at the time t each of the curves has some value, and all these values concentrate very strongly near $(1-2\mu)^t$. If I were to draw the exponential curve $(1-2\mu)^t$, then I will observe that at any fixed t most of the curves lie very close to it. I haven't demonstrated this yet. All I have demonstrated is that the average is exactly $(1-2\mu)^t$. Now I need something slightly more refined, namely the variance. I will skip this calculation -- it's elementary, but it's much more lengthy. What you do is to calculate the average of this thing:

$$\left[\frac{N_b(t)-N_w(t)}{n}-\left(1-2\mu\right)^t\right]^2 \qquad (33).$$

over all possible positions of the set S. This is known as the variance; and if you take the square root you get the standard deviation. If you calculate the standard deviation you will discover (and this is very interesting) that with t fixed as n goes to infinity, the standard deviation is of order $1/\sqrt{n}$. Roughly speaking, using the statistical 3σ business, this means that only a negligible portion of my values lie more than three standard deviations away from the mean $(1-2\mu)^t$. Now remember that n went to infinity, so the larger n is the more these values concentrate at the mean. So really, it takes extremely bad luck to observe a sizable deviation from this $(1-2\mu)^t$.

However, in spite of the convincingness of this argument, one should remember that an arbitrary assumption is floating around. Namely, in performing my average I count all the sets S as being equal -- as being peers of each other. This is something which one cannot justify. It seems natural, because why should one set S be better than another one? But still, you can certainly choose a set S for which the whole thing will deviate very much -- all you have to do is to pick a set S very regularly. But the point that is made obvious is that such a set must be very special. Because they are really proportionately extremely small in number. The predominant number of sets will lead to this nice decreasing exponential curve. And that's how thermodynamics or thermodynamical conclusions ought to be understood. An appropriate way of stating it is that for most of the configurations you're going to observe what you think you ought to observe. Why nature is so kind that it agrees with that, is a philosophical question that I am not prepared to discuss with you here, because I really don't know.

Since we have already learned something from this model, I would like to push it somewhat further and finally to connect directly with the Boltzmann equation and some of the more refined ways of discussiong stochastic phenomena in physics. The first step is to find some way to avoid the integrals. These integrals plague me because I was such an honest fellow and stuck laboriously and religiously to sets with exactly m elements. Now everybody knows that I could allow certain leeway in the number of elements, so I'm going to simplify my life in the following way. I will assume that only on the average does the set S have m elements. The set S may have more or it may have less, but on the average it ought to have m. I did not make this assumption before, because one ought to go through the calculation once, at least, and see that nothing bad happens. Besides you have seen a little trick which may prove useful to you in the future.

But now I'm going to change the model slightly, and assume the following. I am not going to determine my set S by placing m points. Instead I am going to have a coin which has the probability μ of falling heads and the probability $1-\mu$ of falling tails. And at each point of my circle I am going to toss this coin. If it is heads, I'm going to put the point in the set S. If it is tails, I'm not going to put it in the set S. So now the ε's themselves are not any more the well defined quantities of one or minus one whether you're in the set S or not. They now depend on the outcome of the flip of the coin -- or in technical jargon, they become random variables. Each ε_j is -1 with probability μ , and +1 with probability $1-\mu$. Moreover, they are completely independent, because I assume that each time I move from one point to another, I'm going to flip this coin again, independently of the previous toss.

Now, how many elements are there in the set S? Well, that is not a well defined quantity anymore -- it is clearly also a random variable. The number of elements in the set S can be found from the summation of all the ε_p. That won't quite give it to me because the sum is $n-2m$ just as before. So you see that

$$m = \frac{n - \sum_p \varepsilon_p}{2} \tag{34}$$

Now what is the average number of elements in the set S? This is just

$$\langle m \rangle = \frac{1}{2}\left(n - \sum_p \langle \varepsilon_p \rangle\right) \tag{35}$$

and the average $\langle \varepsilon_p \rangle$ is simply $1-2\mu$. So by the time you figure this out, you get

$$\langle m \rangle = n\mu \tag{36}$$

and μ was $\frac{m}{n}$ you remember, so on the average I have the right number of elements. Also you can show that if you take the actual number of elements in S minus the average number, if you square this, average it, and then take the square root to get the standard deviation, then this is of the order of the square root of n. You might say that the number of elements in S is roughly $n\mu$ with an error of the order of $\sqrt{n}$. That means very close to $n\mu$, so we'll expect that a result obtained from this model ought to be exactly the same as the result obtained from the other model. And this one is going to be much easier.

Those of you who have studied a little bit of the standard statistical mechanics may have heard of the grand canonical ensemble. The grand canonical

ensemble is the one in which you allow the number of particles in your system to vary. Yet the results of calculations using the grand canonical ensemble are exactly the same -- because although you allow the number of particles to vary, the mean is the prescribed one and the variations are very small. What I am doing right now, really, is replacing a canonical by a grand canonical ensemble. If you look at the text books on statistical mechanics, you will see that with the canonical ensemble you always have the steepest descent calculation, always these complicated integrals. With the grand canonical ensemble you avoid them. Of course, you have to prove the equivalence, and that is usually done only roughly. Sometimes they aren't equivalent, but I am not going to worry you with that. In simple cases, it stands to reason that if you allow the number of elements to vary, very little, the results should not change very much.

Now, had I chosen this model from the beginning, I'd have no difficulties at all. You remember that I had to calculate

$$\langle \varepsilon_1 \varepsilon_2 \cdots\cdots \varepsilon_t \rangle \quad . \tag{37}$$

But now the averaging is different. We not only have many possible positions of the set S, but now we have also the variability in the number of elements. In fact, the averaging is now simply the kind of averaging you would use in playing this game of tossing a coin. And everything is extremely simple because the tosses are independent; I assumed so. The probability theory that everybody knows tells us that the average of a product of independent things is the product of the averages. Therefore,

$$\langle \varepsilon_1 \varepsilon_2 \cdots\cdots \varepsilon_t \rangle = \langle \varepsilon_1 \rangle \langle \varepsilon_2 \rangle \cdots\cdots \langle \varepsilon_t \rangle \,. \tag{38}$$

Since all the averages are equal, this is simply the average $\langle \varepsilon \rangle^t$. But we have seen that $\langle \varepsilon \rangle = 1-2\mu$, so again you find the result

$$\langle \varepsilon_1 \varepsilon_2 \cdots \varepsilon_t \rangle = (1-2\mu)^t \qquad (39)$$

We are not surprised, but at the same time we ought to be mildly pleased. Since we also got this result from the exact calculation, there clearly is no question of the agreement within the canonical and grand canonical treatments in this particular case. And we have here the wonderful thing that this calculation was immediate. Now with this change in model, I am going to rediscuss the whole problem.

(Remark: I always have to assume that μ is less than 1/2. If it is equal to 1/2, then nothing will ever change from the average because $1-2\mu$ is zero. If μ is bigger than 1/2 it is very interesting, because the whole thing oscillates. We have to have μ less than 1/2 so that the change of color will not be too common an event. That's all the assumption I need, and I will assume from now on that this is so. Of course, if μ is close to 1/2, then you will change colors so actively, and so often, that the equalization will take place very quickly.)

I would like now to tune up the whole problem anew and bring out further analogies with thermodynamics. In this treatment we'll try to parallel that of Gibbs. The idea of Gibbs was the following: Initially, at time t=0, you are given a certain distribution of systems in phase space. Then as time progresses, this distribution evolves in a certain way. You try to prove that in some sense this distribution becomes more and more uniform on the energy surface. To put

it another way, if you have a localized distribution at time t=0 -- the probability is concentrated mainly over a small region of space -- then to begin with you know where your system is, more or less. As time goes on, you know less and less of where the system is. The system, so to speak, wanders around, dissolves. And the knowledge as to where it is gets less and less because of the motion. This is not quite correct, because you really have to introduce all sorts of auxiliary notions such as the coarse-grained density, etc. But one can imitate still, to some extent, what Gibbs tried to do and then see where one runs into difficulties again. Fortunately, for this particular model, one can solve everything. The behavior is simple enough that one can see easily what happens.

The model is the one with the modification that the ε_p are now independent (chosen independently by tossing this μ-coin, you might say). Let us produce a dictionary of terms. First, what is phase space for our model? Well, it's a very simple space in this case. In mechanics, it's the space of all coordinates and momenta -- that means all the attributes required to define the system uniquely. What do I have to do to define my system uniquely? I have to know which ball is black and which is white on each point. That is, I have to know a sequence of plus and minus ones. So the phase space, here, is simply a finite set of points, each point being a sequence of n plus or minus ones. There are 2^n such sequences -- that is, 2^n points in my phase space.

Now suppose that initially at time zero I have a certain distribution $\rho(\vec{\eta}, 0)$ ($\vec{\eta}$ is a point in my phase space, that is a sequence of $\pm$ 1's, which I abbreviate by a vector). $\vec{\eta}$ can assume 2^n different values so that $\rho(\vec{\eta}, 0)$ is really a set of numbers. These numbers give me a distribution if

none are negative, and if they all add up to one. That is, if

$$\rho(\vec{\eta},0) \geq 0 \quad \text{and} \quad \sum_{\vec{\eta}} \rho(\vec{\eta},0) = 1 \tag{40}$$

I have a distribution. For example, you can say that $\rho(\vec{\eta},0)$ is 1 for the vector $\vec{\eta} = (1,1,1,\cdots,1)$, and is 0 for all other ones. That would mean that I would know initially with complete certainty that all the balls are black. But I don't have to take this particular distribution. I can say that I have only partial knowledge at the beginning, then take any distribution I like for $\rho(\vec{\eta},0)$.

The question is, how does the distribution evolve in time? That depends, of course, on the position of the set S. The distribution evolves in time according to the equation

$$\rho(\eta_1,\cdots,\eta_n;t+1) = \rho(\varepsilon_1\eta_2,\varepsilon_2\eta_3,\cdots,\varepsilon_n\eta_1;t) \tag{41}$$

This is the equation of evolution of $\rho(\vec{\eta},t)$. For those of you who are familiar with the terminology, this is the Liouville equation for the model. It tells you how the initial density of an ensemble of systems evolves in time, if the individual systems follow the equations of motion. In statistical mechanics, the Liouville equation unfortunately cannot be solved explicitly. This model, of course, was so designed as to be able to solve it. Let me now decide that since I already have probability in the whole business, I will treat the problem probabilistically.

Let me completely forget that there is a deterministic way in which the whole thing evolves. I can set up what will soon emerge as a very simple

Markov chain and argue as follows. The state of a system is defined entirely by a vector such as $\vec{\eta}$. At each time step I make a transition from this vector to another vector depending on what my set S is. What I know is that if at time t my model is in the state $\vec{\delta}$ then at time t+1 it will be in the state $\vec{\eta}$ where $\vec{\eta}$ is simply related to $\vec{\delta}$. Namely,

$$\begin{aligned} \delta_1 &= \varepsilon_1 \eta_2 \\ \delta_2 &= \varepsilon_2 \eta_3 \\ &\vdots \\ \delta_n &= \varepsilon_n \eta_1 \end{aligned} \tag{42}$$

If I know what my set S is, then I know the ε's -- so I know the exact transition. That's perfectly clear. But in this model, the ε's are random variables. So I must ask, what is the probability of this transition? That is, what is the probability that I make a specified transition in one step. Let me call this transition probability $P(\vec{\delta}\,|\,\vec{\eta})$. It is given by

$$P(\vec{\delta}\,|\,\vec{\eta}) = \text{Prob}\left\{\varepsilon_1 = \delta_1\eta_2\,,\ \varepsilon_2 = \delta_2\eta_3\,,\ \cdots,\ \varepsilon_n = \delta_n\eta_1\right\} \tag{43}$$

(remember that all the ε's , η's , and δ's are either plus one or minus one, so whether you write ε or $\frac{1}{\varepsilon}$, η or $\frac{1}{\eta}$, δ or $\frac{1}{\delta}$, it doesn't matter. It is a very nice algebra). Now this can be easily calculated because I know that the ε's are independent. To say that the epsilons are chosen independently means exactly that

$$\text{Prob}\left\{\varepsilon_1 = \delta_1\eta_2\,,\ \varepsilon_2 = \delta_2\eta_3\,,\ \cdots\cdots,\ \varepsilon_n = \delta_n\eta_1\right\} = \tag{44}$$

$$\prod_{k=1}^{n} \text{Prob}\left\{\varepsilon_k = \delta_k\eta_{k+1}\right\}$$

Now there comes a little bit of a problem --- a very minor one. I don't know whether $\delta_k \eta_{k+1}$ is plus one or minus one. It can be either. So the best way to write it is

$$\text{Prob}\left\{\varepsilon_k = \delta_k \eta_{k+1}\right\} = \frac{1}{2} + \frac{1-2\mu}{2}\,\delta_k \eta_{k+1} \tag{44'}$$

Because now if $\delta_k \eta_{k+1}$ is equal to plus one, then I'm going to get $1-\mu$ for the probability that ε_k is one. If $\delta_k \eta_{k+1}$ is minus one, then I am going to get μ exactly, so this works very nicely. So now the transition probability is just

$$P(\vec{\delta}\,|\,\vec{\eta}) = \prod_{k=1}^{n} \left\{\frac{1}{2} + \frac{1-2\mu}{2}\,\delta_k \eta_{k+1}\right\} \tag{43'}$$

Now that I have the probability of a transition from a state $\vec{\delta}$ to a state $\vec{\eta}$ I'm really in a position to get somewhere. Because now I can make use of a very well developed theory -- the theory of Markov chains.

Now remember what the initial distribution is. It's 2^n positive numbers which add up to one. I can think of the initial distribution as a vector with 2^n components. In a similar way, I can think of the transition probability $P(\vec{\delta}\,|\,\vec{\eta})$ as a matrix of order $2^n \times 2^n$. To see what the distribution after one step is we just multiply the vector $\vec{\rho}(0)$ (that is, the initial distribution) by the matrix. To see what happens after two steps we multiply again by the same matrix, which I will call P. Consequently, to see what happens after t steps, we simply need to raise the matrix to the power t:

$$\vec{\rho}(t) = P^t \vec{\rho}(0) \tag{45}$$

So, if I accept this theory, then all I have to do is find P^t and apply it to the initial distribution. This gives me the distribution at time t. In particular, what I want to show is that no matter where I start at the beginning -- that is, no matter what $\vec{\rho}(0)$ I choose -- if t becomes large enough, then the components of the vector $\vec{\rho}(t)$ approach the same values. So if I am allowed to replace my model by a Markov chain, then everything will depend on the properties of the matrix P. I hope that with a certain amount of cleverness I will be able to decipher these properties.

This method of approach is now known in the trade as the Master equation approach. I will tell you why it is called that in a moment. If you simply write what happens in the transition from time t to time t plus 1:

$$\vec{\rho}(t+1) = P\vec{\rho}(t) \tag{46}$$

then this is known as the "Master equation." This term was proposed by Uhlenbeck many years ago and it was thought initially that the "Master" referred to him. But the "Master" in this terminology refers to this equation which gives you all the information you need. For instance, in the equation we had before, where I considered the number of black balls minus the number of white balls, the information was only partial. Even though I knew how the excess of black balls was going to behave, I didn't really know what the probability was that a black ball sits in place 17 after a time t. But if I am so incredibly insistent on having such information, I can decipher it from this Master equation because it gives me precisely, at each time, the probability of any given situation. That's why it's called the Master equation.

Now the ordinary procedure in physics is to simply assume the Master equation. You say that between every pair of states I have a possible transition.

Somehow I discover the transition probability, and then from there on I just study this equation. It is quite interesting pedagogically -- and I will show it to you, although I will have to skip many details -- that I can derive such an equation for a gas. And then, from this equation, which is linear, derive the non-linear equation of Boltzmann. So it will be an interesting example where a non-linear equation is a consequence of a linear one -- and that is an interesting phenomenon which ought to be looked at a little bit.

But the one remaining problem now facing many people -- and a great deal of investigation is being done on it -- is whether or not this approach can be justified. On our example we can test whether it is justifiable or not. Notice what I have done. I get only the transition probability for one step in time. I then use the resulting matrix to propagate the distribution forever. But in reality, what I should do is solve the Liouville equation (equation 41) and then do the averaging at the end. In both cases, I am averaging over the positions of the set S. Only in one case I do the averaging and then the propagation, and in the other the propagation and then the averaging. In fact, for those of you who are familiar with the terminology, the first procedure which goes with the Master equation is called repeated averaging. Whereas, what really makes sense, is to solve the Liouville equation for time t and then perform the averaging. Now it's not at all clear that these things are interchangeable, that these procedures are equivalent. Because notice what one really does. One can put it picturesquely, although not quite correctly, as follows. The one relation between $\rho(t)$ and $\rho(0)$ we can symbolize by

$$\rho(t) = \langle L\, \rho(t-1) \rangle = \langle L \rangle \rho(t-1) = \langle L \rangle^t \rho(0) \qquad (47)$$

You get the propagation by repeated averaging. The other corresponding to

$$\rho(t+1) = \langle L^t \rho(0) \rangle = \langle L^t \rangle \rho(0) \qquad (47')$$

would be the modified Gibbs approach, to solve the Liouville equation if you can (you almost never can) and then average.

People usually manage to give at least a justification, but not quite a proof, that the results are essentially the same for the dynamical system. It's a very profound fact, because it makes life so much simpler. Because you don't have to solve Liouville's equation, which is tantamount to solving the equations of motion. So if you could once and for all prove that it doesn't matter if you'd average the t^{th} power of the operator of if you average the operator and raise it to the t^{th} power, then you'd be in very good shape. As I say, in most of the present day applications of probability, statistical mechanics, and kinetic theory everyone always makes this assumption anyway. He takes what happens during a very short time, averages the transition operator over that short time, and uses this average operator as the one which propagates. That's done not only in classical statistical mechanics, but in quantum statistical mechanics as well. For instance, in the supposed derivation of the transport equation, first done by Pauli and Fierz. You always say, well there are some random phases; and after a short time you average over them. Then you go for a little time and you again average over them. You constantly keep averaging, whereas in reality you ought to wait until the end of the calculation.

Now in the next lecture I will show you first of all how one can write exactly the solution for both of these approaches. Then we will see how the two things differ, and try to talk ourselves out of this difference. Then I will show you, with some detail, how a corresponding Master equation can be written for an ideal gas with binary collisions. The Master equation will be a linear equation from which all general conclusions can be derived: the H-theorem and various

other things. I will show you how it is related to the non-linear Boltzmann equation which was discussed with you by Professor Dresden. And then with this background we can go on and look at some other stochastic models which, although different, are not unrelated. For instance, in the theory of Brownian motion a very similar treatment can be done. The operators are more complicated and the Master equation becomes a diffusion equation. But, still, fundamentally the same ideas persist.