Generalization of the Ehrenfest Urn Model*

MARTIN J. KLEIN
Department of Physics, Case Institute of Technology, Cleveland, Ohio
(Received February 15, 1956)

A generalization of the usual Ehrenfest urn model is discussed. The generalization is made by relaxing the condition that the transition probabilities between the two urns are equal. The model is interpreted as a schematic representation of the behavior of a system in contact with a heat bath at fixed temperature. The stochastic equations for the model are solved by methods used by Kac and the results are interpreted physically.

I. INTRODUCTION

THE Ehrenfest urn model[1] has proved to be a suggestive guide to the behavior of a closed physical system. In this note we discuss a generalization of the Ehrenfest model which represents schematically the behavior of a system in contact with a heat bath. The equations for this generalized model have been solved by procedures closely analogous to those used by Kac[2] for dealing with the usual model. There are some novel features in the results which are discussed in the final section of this paper.

II. MODEL AND ITS STATIONARY STATE

We consider two urns, A and B, and $2R$ balls numbered consecutively from 1 to $2R$. An integer between 1 and $2R$ is chosen at random in such a way that each of these integers has an equal probability of being chosen. If the ball whose number corresponds to the integer chosen is in A, then we toss a loaded coin α. If the coin turns up "heads," an event whose probability is p, we move the ball in question to urn B; if the toss results in "tails," an event whose probability is $(1-p)$ or q, then the ball is left in urn A. If, however, the ball with the designated number is in urn B, then we toss a different loaded coin β whose fall will move the ball to A with probability p' or leave it in B with probability $(1-p')$ or q'. This whole procedure is repeated regularly at intervals of time τ.

It is evident that if p and p' are equal to one we have the usual case, and this fact gives us a useful check on our results.

We can get a more "physical" picture of our system if we change the terminology.[3] Consider first the usual Ehrenfest model. We think of the two urns as two energy states and the $2R$ balls as $2R$ particles which can occupy these two states. Our basic process is then the transition of one particle from state A to state B (or the reverse) at regular times separated by equal intervals τ. When there are $(R+l)$ particles in A and $(R-l)$ particles in B the probability is $(R+l)/2R$ that the transition is from A to B and the probability is $(R-l)/2R$ that the transition is in the reverse direction. Since individual transitions in both directions are equally likely, we say that our model represents an isolated system and states A and B have the same energy.

In our generalized system, only the transition probabilities are changed. At each of the times when an event can occur, the probability is $p(R+l)/2R$ that the event is an A to B transition, and the probability is $p'(R-l)/2R$ that the transition is in the reverse direction. There is now the additional possibility that no transition occurs and its probability is the difference between one and the sum of the two probabilities just given. Since the individual transitions in the two directions are no longer equally probable we say that our system acts as though the states A and B differ in energy by an amount ϵ and the system is in contact with a heat bath at temperature T.[4] The ratio ϵ/kT (where k is Boltzmann's constant) is defined by the equation

$$p/p' = \exp(-\epsilon/kT). \tag{1}$$

We have chosen ϵ to be the energy of state B on a scale where the energy of state A is zero. We assume for definiteness that ϵ is positive.

We can now write the basic stochastic equation for the system. Let $P(n|m;s)$ be the probability that at time $s\tau$ state A contains $(R+m)$ particles given that A contained $(R+n)$ particles initially. Then $P(n|m;s)$ must be related to the corresponding probabilities at time $(s-1)\tau$ by the equation

$$\begin{aligned} P(n|m;s) &= p\frac{R+m+1}{2R}P(n|m+1;s-1) \\ &\quad + p'\frac{R-m+1}{2R}P(n|m-1;s-1) \\ &\quad + \left(q\frac{R+m}{2R}+q'\frac{R-m}{2R}\right)P(n|m;s-1); \end{aligned} \tag{2}$$

* This research was supported by a grant from the National Science Foundation.

[1] P. Ehrenfest and T. Ehrenfest, Physik. Z. **8**, 311 (1907).

[2] M. Kac, Am. Math. Monthly **54**, 369 (1947). This paper is reprinted in N. Wax, *Selected Papers on Noise and Stochastic Processes* (Dover Publications, Inc., New York, 1954).

[3] A. J. F. Siegert, Phys. Rev. **76**, 1708 (1949). Siegert actually considers a generalization analogous to ours but for the case where the time variable is continuous.

[4] M. J. Klein and P. H. E. Meijer, Phys. Rev. **96**, 250 (1954).

the initial condition is that $P(n|m;\,0)$ is equal to δ_{nm}.

Before we solve Eq. (2) we shall determine the stationary state of the system. In this stationary state the probability of finding $(R+m)$ particles in state A is given by $P_0(m)$, where $P_0(m)$ satisfies the equation

$$P_0(m)=p\frac{R+m+1}{2R}P_0(m+1)+p'\frac{R-m+1}{2R}P_0(m-1)+\left(q\frac{R+m}{2R}+q'\frac{R-m}{2R}\right)P_0(m);\quad (3)$$

in other words the distribution $P_0(m)$ is invariant in time. It is easily verified that the solution of Eq. (3) is the $P_0(m)$ given by the expression

$$P_0(m)=\frac{(2R)!}{(R+m)!(R-m)!}\left(\frac{p}{p'}\right)^{R-m}\frac{1}{(1+p/p')^{2R}}.\quad (4)$$

We notice that when p and p' are equal, whether or not they are equal to one, this expression reduces to the stationary distribution for the usual Ehrenfest model. If we introduce Eq. (1) we can rewrite $P_0(m)$ in the form

$$P_0(m)=\frac{(2R)!}{(R+m)!(R-m)!}\,\frac{\exp\{-(R-m)\epsilon/kT\}}{\{1+\exp(-\epsilon/kT)\}^{2R}}.\quad (5)$$

This last expression will be recognized at once as the probability of finding $(R-m)$ particles in energy state ϵ when the system (at temperature T) contains $2R$ particles each of which can have energy zero or ϵ. Our stationary distribution is therefore consistent with the physical interpretation for the model given above.

III. SOLUTION OF THE STOCHASTIC EQUATION

The method used to solve Eq. (2) is that used by Kac[2] in his solution of the corresponding equation for the Ehrenfest model. The key idea is to consider the set of numbers $P(n|m;s)$ for m running from $-R$ to R, as a vector $P(s)$ and to write Eq. (2) schematically as

$$P(s)=AP(s-1),\quad (6)$$

where A is a nonsymmetric matrix whose elements are determined by Eq. (2). Introduce the eigenvalues λ, right eigenvectors $\boldsymbol{\xi}$ and left eigenvectors $\boldsymbol{\eta}$ of A by the equations

$$A\boldsymbol{\xi}=\lambda\boldsymbol{\xi}\quad (7)$$

and

$$\tilde{A}\boldsymbol{\eta}=\lambda\boldsymbol{\eta},\quad (8)$$

where $\tilde{A}$ is the transpose of A. Now let Ξ be the matrix whose columns are the vectors $\boldsymbol{\xi}$, H be the matrix whose rows are the vectors $\boldsymbol{\eta}$, and Λ be the diagonal matrix whose diagonal elements are the eigenvalues λ. Then since A is given by the product $\Xi\Lambda\mathrm{H}$ due to the orthogonality of $\boldsymbol{\xi}$'s and $\boldsymbol{\eta}$'s belonging to different eigenvalues, it follows that the solution to the basic equation is given by the expression

$$P(s)=A^sP(0)=\Xi\Lambda^s\mathrm{H}P(0).\quad (9)$$

Kac's method for determining the eigenvalues and eigenvectors can also be applied to our system. If we denote the components of the $\boldsymbol{\xi}$ vector by x_k, where k runs from 0 to $2R$, then the set of linear equations satisfied by the x_k is

$$\begin{aligned}
&q'x_0+p\frac{1}{2R}x_1=\lambda x_0,\\
&p'x_0+\left(q\frac{1}{2R}+q'\frac{2R-1}{2R}\right)x_1+p\frac{2}{2R}x_2=\lambda x_1,\\
&\cdots\\
&p'\frac{2R-(k-1)}{2R}x_{k-1}+\left(q\frac{k}{2R}+q'\frac{2R-k}{2R}\right)x_k+p\frac{k+1}{2R}x_{k+1}=\lambda x_k,\\
&\cdots\\
&p'x_{2R-1}+qx_{2R}=\lambda x_{2R}.
\end{aligned}\quad (10)$$

It is convenient to replace this set of $(2R+1)$ equations by the infinite set in which the first $2R$ equations are the same, and in which the $(2R+1)$th equation has, on its left hand side, the added term $p[(2R+1)/2R]x_{2R+1}$. The equations for k equal to $2R+l$ are then the same as the equations for k equal to $l-1$, for all positive integers l. Of course we must find solutions for which x_{2R+1} is zero, and therefore all higher x_k vanish as well.

Multiply the equations by z^k and sum. We then obtain, after some simple manipulation, a differential equation for the function $f(z)$ which is defined by the relation

$$f(z)\equiv\sum_{k=0}^{\infty}x_kz^k.\quad (11)$$

The differential equation is

$$f'(z)\left[\frac{p}{2R}-\frac{p'}{2R}z^2+\frac{q-q'}{2R}z\right]=f(z)[\lambda-p'z-q].\quad (12)$$

Equation (12) can be solved directly and the solution is given by the expression

$$f(z)=c(p+p'z)^{2R-\sigma}(p'-p'z)^{\sigma},\quad (13)$$

where c is a constant and σ is an abbreviation for $2R(1-\lambda)/(p+p')$. The requirement that $f(z)$ be a polynomial of degree $2R$ means that σ must be an integer between zero and $2R$. This requires that λ must be

given by the expression

$$\lambda_j=[j(p+p')/2R]+1-(p+p')/2, \tag{14}$$

where j is any integer in the set $-R$ to R. We see that there are precisely $2R+1$ distinct eigenvalues, and that the largest is equal to one. Introducing this expression for λ and making use of the condition that $f(0)$ should be equal to x_0 leads us to a final form for $f(z)$ (and therefore for the components of the right eigenvectors) given by

$$f(z)=x_0[1+(zp'/p)]^{R+j}(1-z)^{R-j}. \tag{15}$$

The left eigenvectors can also be found by Kac's methods. We omit the derivation and proceed directly to the final result for $P(n|m;s)$. As a matter of notation we let $D_k^{(j)}(p'/p)$ be the coefficient of z^k in $[1+(zp'/p)]^{R+j}(1-z)^{R-j}$, where j takes on the integer values from $-R$ to R and k takes on the integer values from zero to $2R$. Then $P(n|m;s)$ is given by the equation

$$P(n|m;s)=\frac{(-1)^{R+m}}{[1+(p'/p)]^{2R}}\left(\frac{p'}{p}\right)^{R-n} \times\sum_{j=-R}^{R} D_{R+m}{}^{(j)}(p'/p)(\lambda_j)^s D_{R+j}{}^{(-n)}(p/p'). \tag{16}$$

[Note that the last factor is $D_{R+j}{}^{(-n)}(p/p')$ since the left eigenvectors involve the ratio p/p' rather than p'/p.]

We have two checks on Eq. (16). First, if p and p' are equal to one then $P(n|m;s)$ must reduce to the expression given by Kac. This may be verified by inspection. Second, as s goes to ∞ we expect that $P(n|m;s)$ will approach the stationary distribution $P_0(m)$ discussed in the previous section. To see this we must notice that all eigenvalues λ_j except λ_R, which is equal to one, are less than one in absolute value. Consequently the sum on j reduces to a single term in the limit, and we can write for $P(n|m;\infty)$ the expression

$$P(n|m;\infty)=\frac{(-1)^{R+m}}{[1+(p'/p)]^{2R}}\left(\frac{p'}{p}\right)^{R-n} \times D_{R+m}{}^{(R)}(p'/p)D_{2R}{}^{(-n)}(p/p'). \tag{17}$$

Now $D_{R+m}{}^{(R)}(p'/p)$ is the coefficient of z^{R+m} in $[1+(zp'/p)]^{2R}$ and therefore is equal to $(p'/p)^{R+m}\times[(2R)!/(R+m)!(R-m)!]$. Similarly, $D_{2R}{}^{(-n)}(p/p')$ is the coefficient of z^{2R} in $[1+(zp/p')]^{R-n}(1-z)^{R+n}$ and therefore is equal to $(-1)^{R+n}(p/p')^{R-n}$. If we substitute these values in Eq. (17), we find that

$$P(n|m;\infty)=\frac{(2R)!}{(R+m)!(R-m)!}\left(\frac{p'}{p}\right)^{R+m} \times\frac{1}{[1+(p'/p)]^{2R}}=P_0(m). \tag{18}$$

[This form for $P_0(m)$ is slightly different from the one in Eq. (14), but if the latter is multiplied by $(p'/p)^{2R}$ in numerator and denominator it is converted into the expression of Eq. (18).]

IV. DISCUSSION

One of the aspects of the behavior of this model which is of interest is the time dependence of $\langle m\rangle$, the mean value of m, i.e., the time variation of the average number of particles in state A. The value of $\langle m\rangle$ is computed from the equation

$$\langle m\rangle=\sum_{m'=-R}^{R} m'P(n|m';s), \tag{19}$$

where $\langle m\rangle$ is actually the mean value of m at time $s\tau$ when m was equal to n at time zero. This expression is evaluated with the help of Eq. (16). We find that

$$\langle m\rangle=\frac{(-1)^{R+n}}{[1+(p'/p)]^{2R}}\left(\frac{p'}{p}\right)^{R-n}\sum_{j=-R}^{R} D_{R+j}{}^{(-n)}(p/p')(\lambda_j)^s \times\sum_{m'=-R}^{R} m'D_{R+m'}{}^{(j)}(p'/p). \tag{20}$$

The sum on m' is evaluated by noticing that

$$\sum_{m'=-R}^{R} m'D_{R+m'}{}^{(j)}(p'/p) =\left\{z\frac{d}{dz}[z^{-R}[1+(zp'/p)]^{R+j}(1-z)^{R-j}]\right\}_{z=1}. \tag{21}$$

A straightforward calculation gives as the final result

$$\langle m\rangle=R\frac{1-(p/p')}{1+(p/p')}+\left\{\frac{(R+n)(p/p')-(R-n)}{1+(p/p')}\right\} \times\left\{\left(1-\frac{p+p'}{2R}\right)^s\right\}. \tag{22}$$

As before we have two ready checks on this equation. When p and p' are equal to one, Eq. (22) reduces to the known equation for the Ehrenfest model.[5] Alternatively, when s goes to infinity $\langle m\rangle$ approaches $R[1-(p/p')]/[1+(p/p')]$. This last expression is, of course, the value of $\langle m\rangle$ in the equilibrium state. We can see the meaning of this last expression most easily by noticing that it implies that the mean number of particles in state B (the upper energy state) is, at equilibrium, just equal to

$$R-\langle m\rangle=2R\exp(-\epsilon/kT)/[1+\exp(-\epsilon/kT)] \tag{23}$$

as expected.

---

[5] See Kac (reference 2), and also E. Schrödinger and K. W. F. Kohlrausch, Physik. Z. **27**, 306 (1926) and B. Friedman, Comm. Pure Appl. Math. **2**, 59 (1949).

We notice that according to Eq. (22) the initial deviation of $\langle m\rangle$ from its value in the equilibrium distribution decays to zero with a "relaxation time" given by $\tau 2R/(p+p')$.

As our final point we consider the time variation of the thermodynamic functions of our model. Since we are dealing with a system which is not isolated but is instead coupled to a heat bath at temperature T, we do not expect the entropy of the system itself to be a maximum at equilibrium. Instead, the entropy of the system plus that of the heat bath must be maximum, or, equivalently, the Helmholtz function of the system must be a minimum at equilibrium. Without giving the details, we mention that it can be proved by the methods we used in another paper,[6] that the Helmholtz function for this system decreases monotonically and attains its minimum value in the equilibrium state. Consequently in this respect, as in all others considered, the generalized Ehrenfest urn model is indicative of the behavior of a system which is kept at a fixed temperature.

[6] M. J. Klein, Physica (to be published).

PHYSICAL REVIEW VOLUME 103, NUMBER 1 JULY 1, 1956

Thermodynamics and Statistical Mechanics at Negative Absolute Temperatures

NORMAN F. RAMSEY*
Harvard University, Cambridge, Massachusetts, and Clarendon Laboratory, Oxford, England
(Received March 26, 1956)

The circumstances under which negative absolute temperatures can occur are discussed, and principles of thermodynamics and statistical mechanics at negative temperatures are developed. If the entropy of a thermodynamic system is not a monotonically increasing function of its internal energy, it possesses a negative temperature whenever $(\partial S/\partial U)_X$ is negative. Negative temperatures are hotter than positive temperatures. When account is taken of the possibility of negative temperatures, various modifications of conventional thermodynamics statements are required. For example, heat *can* be extracted from a negative-temperature reservoir with no other effect than the performance of an equivalent amount of work. One of the standard formulations of the second law of thermodynamics must be altered to the following: It is impossible to construct an engine that will operate in a closed cycle and provide no effect other than (1) the extraction of heat from a positive-temperature reservoir with the performance of an equivalent amount of work or (2) the rejection of heat into a negative-temperature reservoir with the corresponding work being done on the engine. A thermodynamic system that is in internal thermodynamic equilibrium, that is otherwise essentially isolated, and that has an energetic upper limit to its allowed states can possess a negative temperature. The statistical mechanics of such a system are discussed and the results are applied to nuclear spin systems.

I. INTRODUCTION

SEVERAL years ago Pound,[1–3] Purcell,[3] and Ramsey[2,4] studied experimentally various properties of the nuclear spin systems in a pure LiF crystal for which spin lattice relaxation times were as large as 5 minutes at room temperature while the spin-spin relaxation time was less than 10^{-5} second. With the nuclear spin systems of this crystal various experiments were carried out, including experiments with a spin system at negative temperatures.[3] In the present paper, the thermodynamical and statistical mechanical implications of negative absolute temperature are discussed. Since the theoretical analysis of the past experiments has been only briefly described,[1–4] there has been some misunderstanding[5] of them. For this reason and because of the thermodynamic significance of negative temperatures,[4,6] the present paper also contains a more detailed justification for the use of temperature as a description of suitable nuclear spin systems.

As discussed in Sec. III below, the conditions for the existence of a system at negative temperatures are so restrictive that they are rarely met in practice except with some mutually interacting nuclear spin systems. However, the thermodynamics and statistical mechanics of negative temperatures are more general than their application to a single type of system. Consequently, in the present paper, the thermodynamics and statistical mechanics of negative temperatures will be discussed first for a general system capable of negative temperatures, and only later will specific applications be made to spin systems.

* John Simon Guggenheim Fellow.
[1] R. V. Pound, Phys. Rev. **81**, 156 (1951).
[2] N. F. Ramsey and R. V. Pound, Phys. Rev. **81**, 278 (1951).
[3] E. M. Purcell and R. V. Pound, Phys. Rev. **81**, 279 (1951).
[4] N. F. Ramsey, Ordnance **40**, 898 (1956).
[5] W. F. Giauque, J. Am. Chem. Soc. **76**, 5577 (1954); Time **46**, No. 26, 49 (1955).
[6] F. J. Dyson, Sci. American (September, 1954); F. Simon (private communication).

II. THERMODYNAMICS AT NEGATIVE TEMPERATURES

From a thermodynamic point of view, the only requirement for the existence of a negative temperature is that the entropy S should not be restricted to a monotonically increasing function of the internal energy U. At any point for which the slope of the entropy as a