PHYSICAL REVIEW VOLUME 96, NUMBER 2 OCTOBER 15, 1954

Principle of Minimum Entropy Production*

MARTIN J. KLEIN† AND PAUL H. E. MEIJER‡
Department of Physics, Case Institute of Technology, Cleveland, Ohio

(Received April 1, 1954)

The principle of minimum entropy production says that the steady state of an irreversible process, i.e., the state in which the thermodynamic variables are independent of the time, is characterized by a minimum value of the rate of entropy production. This theorem, due to Prigogine, is proved by the methods of statistical mechanics for a particular process—the flow of matter and energy through a narrow tube connecting two containers of an ideal gas. The two containers are maintained at slightly different temperatures. The resultant form for the entropy production in the steady state, and the method used in the proof, give additional insight into the significance of the principle of minimum entropy production.

I. INTRODUCTION

THE principle of minimum entropy production gives a thermodynamic criterion for determining the steady state of a system in which an irreversible process is taking place. The criterion can be formulated very simply: the steady state is that state in which the rate of entropy production has the minimum value consistent with the external constraints which prevent the system from reaching equilibrium. The equilibrium state takes its natural place in this way of looking at thermodynamic processes as the simplest kind of steady state. When there are no constraints, the system proceeds to that state in which the rate of entropy production is zero, i.e., to the equilibrium state. When constraints, such as an externally maintained temperature difference between parts of the system, prevent the system from reaching equilibrium, the system does the next best thing: it goes to a state in which the thermodynamic variables do not change with time and in which the smallest possible amount of entropy is created per unit time.

This characterization of the steady state as the state of minimum entropy production is due to Prigogine[1] and, in a somewhat generalized form, to de Groot.[2] The proofs given by both authors are based on the Onsager[3] relations, which form the starting point of the recently developed thermodynamic theory of irreversible processes in the steady state.[4]

Because the principle of minimum entropy production has a wide range of applications and because it seems to pick out an essential feature of irreversible processes, we have tried to derive it by the methods of statistical mechanics. The method we have used is a generalization of the method used by Pauli[5] to derive the second law of thermodynamics (or H theorem) from quantum mechanics. This method has recently been discussed by Thomsen,[6] and we have drawn upon Thomsen's paper in our own work.

Using this method we have derived the principle of minimum entropy production for a particular irreversible process—the flow of matter and energy through a small capillary connecting two containers of an ideal gas. The two identical containers are maintained at slightly different temperatures by two heat baths. Starting with a set of equations which gives the time variation of the numbers of molecules in the energy states of the two containers, we have proved that the steady state corresponds to that set of occupation numbers which minimizes the rate of entropy production. Two features of interest in our derivation are the necessity for a small temperature difference between the two containers and the important role played by the two heat baths in the entropy production. Our final form for the rate of entropy production in the steady state has a simple physical interpretation relating the entropy production to the flow of heat from one heat bath to the other.

We conclude this introduction with an outline of the contents of the subsequent sections. Section II contains the detailed specification of the system considered, and includes the basic stochastic equations for rates of change of the relevant probabilities. Section III contains the derivation of the equations for the rate of production of entropy, and includes a discussion of the role played by the heat baths. Section IV contains the proof that the steady state is the state of minimum entropy production, and includes a discussion of the necessary assumptions. The final section contains a brief discussion of the entropy production in the steady state.

II. SPECIFICATION OF THE SYSTEM

Our physical system consists of two identical containers, C_1 and C_2, each of volume V, containing a total

---

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

† This work was begun during the tenure of a National Research Fellowship at the Dublin Institute for Advanced Studies.

‡ Travel from the Netherlands was made possible by a Fulbright travel grant.

[1] I. Prigogine, *Etude Thermodynamique des Phênomènes irréversibles* (Editions Desoer, Liège, 1947), Chap. V.

[2] S. R. de Groot, *Thermodynamics of Irreversible Processes* (Interscience Publishers, Inc., New York, 1951), Chap. X.

[3] L. Onsager, Phys. Rev. **37**, 405 (1931); **38**, 2265 (1931).

[4] See references 1 and 2, and also K. G. Denbigh, *The Thermodynamics of the Steady State* (Methuen and Company, Ltd., London, 1951), and L. Onsager and S. Machlup, Phys. Rev. **91**, 1505, 1512 (1953).

[5] W. Pauli in *Probleme der Modernen Physik*, edited by P. Debye (S. Hirzel, Leipzig, 1928), p. 30.

[6] J. S. Thomsen, Phys. Rev. **91**, 1263 (1953).

of N molecules of an ideal gas. The two containers are connected through a narrow capillary tube.[7] We assume that the dimensions of the capillary tube are so chosen that collisions between molecules in the tube can be neglected. We assume also that collisions of the molecules with the walls of the tube do not affect the energies of these molecules. Each container is in good thermal contact with a heat bath, the heat baths having temperatures T_1 and T_2, respectively. We shall eventually have to assume that T_1 and T_2 differ by an amount small compared to either, but we shall not make this assumption until it becomes necessary.

Let us denote the energy levels of a molecule in one of the boxes of volume V by ϵ_i $(i=1, 2, \cdots)$. Since the boxes are identical the energy levels are the same in both containers. We can specify the state of the system at any time by giving the values of the two sets of occupation numbers $n_i^{(1)}$, $n_i^{(2)}$ at that time. The symbol $n_i^{(1)}$ represents the number of molecules in the ith energy state in C_1, and $n_i^{(2)}$ is defined similarly.

For later developments it is slightly more convenient to work in terms of the occupation probabilities rather than the occupation numbers. We, therefore, make the following definitions:

$$p_i=n_i^{(1)}/N, \quad q_i=n_i^{(2)}/N. \tag{1}$$

We notice that p_i is the probability of finding a molecule in the ith energy state of C_1, and q_i is defined similarly. The p_i and q_i are subject to the obvious restriction

$$\sum_i (p_i+q_i)=1, \tag{2}$$

which expresses the constancy of the total number of molecules.

Our first step is to state the equations giving the time variation of the p_i and q_i. Let us define a_{ij} as the probability per unit time that a molecule in C_1, originally in state i makes a transition to state j, remaining in C_1. We define c_{ij} as the corresponding transition probability per unit time for a molecule in C_2. Let b_{ii} be the probability per unit time that a molecule in state i moves from C_1 to C_2 or conversely. By the assumption discussed above we need not consider any but the diagonal b_{ii}.

The equations[8] for the time rates of change dp_i/dt and dq_i/dt are now taken to be

$$dp_i/dt=\sum_j (a_{ji}p_j-a_{ij}p_i)+b_{ii}(q_i-p_i), \tag{3}$$

and

$$dq_i/dt=\sum_j (c_{ji}q_j-c_{ij}q_i)+b_{ii}(p_i-q_i). \tag{4}$$

Equation (3) can readily be interpreted: the time rate of change of p_i is the sum of the probabilities that molecules come into state i (in C_1) from all other states j (in C_1), plus the probability that molecules come into state i from $C_2(b_{ii}q_i)$ minus the corresponding probabilities that molecules leave state i for the other states in C_1 or for C_2.

We must now point out the relations among the a_{ij} and among the c_{ij} which follow from the fact that C_1 and C_2 are in contact with heat baths at temperatures T_1 and T_2, respectively. The relations in question are

$$a_{ij}e^{-\epsilon_i/kT_1}=a_{ji}e^{-\epsilon_j/kT_1}, \tag{5}$$

and

$$c_{ij}e^{-\epsilon_i/kT_2}=c_{ji}e^{-\epsilon_j/kT_2}. \tag{6}$$

These equations are derived in the Appendix. The derivation is based upon the conservation of energy for the system of container and heat bath, and upon the rapid increase of the density of energy states of a macroscopic system with increasing energy.

III. THE ENTROPY PRODUCTION

The rate of entropy production in our problem can now be expressed in the following schematic form:

$$dS/dt=(dS/dt)_{C1}+(dS/dt)_{C2} +(dS/dt)_{B1}+(dS/dt)_{B2}. \tag{7}$$

The first pair of terms express the rates of entropy production in the gas in the two containers, and the second pair of terms express the entropy production in the two heat baths. The presence of this second pair of terms is worth special mention; we must include them since the heat baths play an important part in the production of entropy. We shall return to a discussion of this point later in the section.

We must now evaluate the four terms of Eq. (7). Consider first the entropy production in the gas, the first pair of terms. The entropy of the gas can be expressed in terms of the molecular occupation numbers as[9]

$$S_{C1}+S_{C2}=-k\sum_i [n_i^{(1)}\log n_i^{(1)}+n_i^{(2)}\log n_i^{(2)}], \tag{8}$$

or in terms of the occupation probabilities as

$$S_{C1}+S_{C2}=-Nk\sum_i [p_i\log p_i+q_i\log q_i]+Nk\log N. \tag{9}$$

Differentiating with respect to the time and substituting from Eqs. (3)–(4), we obtain

$$\begin{aligned}(dS/dt)_{C1}+(dS/dt)_{C2} &=-Nk\Big[\sum_i \log p_i\Big\{\sum_j (a_{ji}p_j-a_{ij}p_i)+b_{ii}(q_i-p_i)\Big\}\\ &\quad+\sum_i \log q_i\Big\{\sum_j (c_{ji}q_j-c_{ij}q_i)+b_{ii}(p_i-q_i)\Big\}\Big]. \end{aligned}\tag{10}$$

---

[7] See the discussion in reference 2, and also in H. B. G. Casimir, Revs. Modern Phys. **17**, 343 (1945).

[8] A. J. F. Siegert, Phys. Rev. **76**, 1708 (1949).

[9] R. C. Tolman, *The Principles of Statistical Mechanics* (Oxford University Press, London, 1938), Chap. VI.

Our next task is the evaluation of the entropy production in the two heat baths. Consider the bath B_1 at constant temperature T_1. We can write, for its rate of entropy production,

$$(dS/dt)_{B1}=(1/T_1)(dU/dt)_{B1}, \tag{11}$$

where U_{B1} is the internal energy of the bath, and where the right-hand side expresses the fact that any energy which enters or leaves the bath does so at temperature T_1. Since the bath B_1 exchanges energy only with the gas in C_1, it follows that the rate of change of U_{B1} is the negative of the rate of change of U_1, the internal energy of the gas in C_1, if we omit from the latter the energy exchanged with container C_2. Thus, we can write the following equation for $(dS/dt)_{B1}$,

$$(dS/dt)_{B1}=-(1/T_1)(dU_1/dt)_{\text{internal}}. \tag{12}$$

Since U_1 is just $N\sum_i \epsilon_i p_i$, we can evaluate the right-hand side by omitting the last term in Eq. (3) and obtain

$$(dS/dt)_{B1}=-(N/T_1)\sum_i \epsilon_i \sum_j (a_{ji}p_j-a_{ij}p_i). \tag{13}$$

Similarly we obtain the equation

$$(dS/dt)_{B2}=-(N/T_2)\sum_i \epsilon_i \sum_j (c_{ji}q_j-c_{ij}q_i). \tag{14}$$

Combining Eqs. (10), (13), (14), we have the final expression for the total rate of entropy production:

$$\begin{aligned}\frac{dS}{dt}=-Nk\sum_i\bigg[\left(\frac{dp_i}{dt}\right)\log p_i+\left(\frac{dq_i}{dt}\right)\log q_i\\ +\frac{\epsilon_i}{kT_1}\left(\frac{dp_i}{dt}\right)_{\text{int.}}+\frac{\epsilon_i}{kT_2}\left(\frac{dq_i}{dt}\right)_{\text{int.}}\bigg]\\ =-Nk\sum_i\bigg[\left(\log p_i+\frac{\epsilon_i}{kT_1}\right)\sum_j(a_{ji}p_j-a_{ij}p_i)\\ +\left(\log q_i+\frac{\epsilon_i}{kT_2}\right)\sum_j(c_{ji}q_j-c_{ij}q_i)\\ +b_{ii}(\log p_i-\log q_i)(q_i-p_i)\bigg].\end{aligned} \tag{15}$$

IV. THE STEADY STATE

We must now prove that the set of values of the p_i and q_i which minimize the rate of entropy production, as given by Eq. (15), are constants in time, i.e., constitute a steady state of the system. Before we formulate our theorem, we shall obtain the equations determining the p_i and q_i which minimize dS/dt.[10]

Differentiating Eq. (15) with respect to p_r we obtain

$$\begin{aligned}\frac{\partial}{\partial p_r}\left(\frac{dS}{dt}\right)=-Nk\bigg[\sum_i\left(\log p_i+\frac{\epsilon_i}{kT_1}\right)(a_{ri}-\delta_{ri}\sum_j a_{rj})\\ +\sum_j\left(a_{jr}\frac{p_j}{p_r}-a_{rj}\right)+b_{rr}\left\{\left(\frac{q_r}{p_r}-1\right)+\log\frac{q_r}{p_r}\right\}\bigg].\end{aligned} \tag{16}$$

Similarly, by differentiating with respect to q_r, we obtain

$$\begin{aligned}\frac{\partial}{\partial q_r}\left(\frac{dS}{dt}\right)=-Nk\bigg[\sum_i\left(\log q_i+\frac{\epsilon_i}{kT_2}\right)(c_{ri}-\delta_{ri}\sum_j c_{rj})\\ +\sum_j\left(c_{jr}\frac{q_j}{q_r}-c_{rj}\right)+b_{rr}\left\{\left(\frac{p_r}{q_r}-1\right)+\log\frac{p_r}{q_r}\right\}\bigg].\end{aligned} \tag{17}$$

We cannot set these derivatives equal to zero because the p_i and q_i are subject to the constraint expressed by Eq. (2). If we introduce a Lagrange multiplier μ to take care of this constraint, and at the same time make use of Eqs. (5) and (6), we can write the conditions for a minimum of dS/dt as

$$\begin{aligned}\mu=Nk\bigg[\sum_i a_{ri}\left\{\log\frac{p_ie^{\epsilon_i/kT_1}}{p_re^{\epsilon_r/kT_1}}+\left(\frac{p_ie^{\epsilon_i/kT_1}}{p_re^{\epsilon_r/kT_1}}-1\right)\right\}\\ +b_{rr}\left\{\log\frac{q_r}{p_r}+\left(\frac{q_r}{p_r}-1\right)\right\}\bigg],\\ \mu=Nk\bigg[\sum_i c_{ri}\left\{\log\frac{q_ie^{\epsilon_i/kT_2}}{q_re^{\epsilon_r/kT_2}}+\left(\frac{q_ie^{\epsilon_i/kT_2}}{q_re^{\epsilon_r/kT_2}}-1\right)\right\}\\ +b_{rr}\left\{\log\frac{p_r}{q_r}+\left(\frac{p_r}{q_r}-1\right)\right\}\bigg].\end{aligned} \tag{18}$$

At this point we must introduce the assumption that the system is not far from equilibrium. We may make this more physical by saying that we require that the temperatures T_1 and T_2 be given by $T+\delta T$ and $T-\delta T$, respectively, where only first order terms in $\delta T/T$ need be maintained. Further, we require that the gases in the two containers be nearly in equilibrium at the respective temperatures T_1 and T_2. We may summarize the restrictions by the equations

$$\frac{q_r}{p_r}=1+\delta_r,\quad \frac{p_je^{\epsilon_j/kT_1}}{p_ie^{\epsilon_i/kT_1}}=1+\rho_{ij},\quad \frac{q_je^{\epsilon_j/kT_2}}{q_ie^{\epsilon_i/kT_2}}=1+\sigma_{ij}, \tag{19}$$

where δ, ρ, and σ are functions of the time in general but are all small, and second-order terms in any of these quantities may be dropped.[11]

Under these assumptions we shall now prove the basic theorem: the necessary and sufficient condition

---

[10] This discussion generalizes that given by Thomsen in reference 6.

[11] Compare with I. Prigogine, Physica **15**, 272 (1949).

that the entropy production is a minimum is that all p_i and q_i are independent of the time, i.e, that the system is in a steady state.

A. Sufficiency

Given that $(dp_i/dt)=0$ and $(dq_i/dt)=0$ (all i), we are to prove that for these values of the p_i and q_i, dS/dt is minimum. From Eq. (3) it follows that

$$b_{ii}(q_i-p_i)+\sum_j(a_{ji}p_j-a_{ij}p_i)=0.$$

If we make use of Eq. (5) this can easily be rewritten in the form

$$b_{ii}\left(1-\frac{q_i}{p_i}\right)=\sum_j a_{ij}\left(\frac{p_je^{\epsilon_j/kT_1}}{p_ie^{\epsilon_i/kT_1}}-1\right). \tag{20}$$

Furthermore, if we make use of the assumption expressed in Eq. (19), we can rewrite Eq. (20) as

$$-b_{ii}\log\frac{q_i}{p_i}=\sum_j a_{ij}\log\frac{p_je^{\epsilon_j/kT_1}}{p_ie^{\epsilon_i/kT_1}}. \tag{21}$$

Now, Eq. (20) and (21) tell us that the equations which determine the minimum for dS/dt, Eq. (18), are satisfied for all r with $\mu=0$. This result, with a similar argument for the terms in the q's, proves the sufficiency of our condition.

B. Necessity

Given that dS/dt is a minimum we are to prove that for these values of the p_i and q_i, dp_i/dt, and dq_i/dt vanish. Making use of Eq. (19) we can rewrite the minimum conditions of Eq. (18) in the form

$$\mu=2Nk\left[\sum_i a_{ri}\left(\frac{p_ie^{\epsilon_i/kT_1}}{p_re^{\epsilon_r/kT_1}}-1\right)+b_{rr}\left(\frac{q_r}{p_r}-1\right)\right]. \tag{22}$$

Multiplying by p_r and using Eq. (5) yields

$$\mu p_r=2Nk(dp_r/dt). \tag{23}$$

Similarly, we can obtain the equation

$$\mu q_r=2Nk(dq_r/dt). \tag{24}$$

If we sum Eq. (23) and (24) over all values of r and take note of Eq. (2) it follows that μ must be zero. Hence all dp_r/dt and dq_r/dt must vanish, establishing our theorem.

One point remains for discussion—the proof that the value of dS/dt is actually a minimum, rather than some other stationary value, in the steady state. To show this we shall compute the change $\delta(dS/dt)$ in the entropy production for small changes δp_i, δq_i in the p_i and q_i from their steady-state values which we denote by p_i^0 and q_i^0. A direct calculation from Eq. (15) gives

$$\delta\left(\frac{dS}{dt}\right)=-Nk\left[\sum_{i,j}\left\{\frac{a_{ji}}{p_i^0}\delta p_i\delta p_j-\frac{a_{ij}}{p_i^0}(\delta p_i)^2+\frac{c_{ji}}{q_i^0}\delta q_i\delta q_j-\frac{c_{ij}}{q_i^0}(\delta q_i)^2\right\}+\sum_i b_{ii}\left\{-\frac{(\delta p_i)^2}{p_i^0}-\frac{(\delta q_i)^2}{q_i^0}+\frac{q_i^0+p_i^0}{p_i^0q_i^0}\delta p_i\delta q_i\right\}\right]. \tag{25}$$

There are no first-order terms because we are computing the change from a stationary value. In order to prove that $\delta(dS/dt)$ is positive we show that it can be expressed as a sum of squares. To do this replace $\sum_j a_{ij}$ and $\sum_j c_{ij}$ by their values from the equations

$$\begin{gathered}\sum_j a_{ji}p_j^0-p_i^0\sum_j a_{ij}+b_{ii}(q_i^0-p_i^0)=0,\\ \sum_j c_{ji}q_j^0-q_i^0\sum_j c_{ij}+b_{ii}(p_i^0-q_i^0)=0,\end{gathered} \tag{26}$$

which express the fact that p_i^0, q_i^0 correspond to a steady state. It is also necessary to add to the resulting expression for $\delta(dS/dt)$ the expression obtained from Eq. (25) by interchanging the summation indices i, j in the second and fourth terms of that equation. We obtain finally

$$2\delta\left(\frac{dS}{dt}\right)=Nk\left[\sum_{i,j}a_{ij}\left(\frac{\delta p_i}{(p_i^0)^{\frac{1}{2}}}-\frac{\delta p_j(p_i^0)^{\frac{1}{2}}}{p_j^0}\right)^2+\sum_{i,j}c_{ij}\left(\frac{\delta q_i}{(q_i^0)^{\frac{1}{2}}}-\frac{\delta q_j(q_i^0)^{\frac{1}{2}}}{q_j^0}\right)^2+\sum_i b_{ii}(p_i^0+q_i^0)\left(\frac{\delta p_i}{p_i^0}-\frac{\delta q_i}{q_i^0}\right)^2\right]\geqq 0, \tag{27}$$

which proves that the stationary value of dS/dt is actually a minimum.

This completes our proof that the steady state is the state of minimum entropy production.

V. DISCUSSION

The minimum entropy production in the steady state can be expressed in an interesting form. If we examine the first form of Eq. (15), we notice that in the steady state the first two terms vanish, and we can write

$$\frac{dS}{dt}=-Nk\sum_i\left\{\frac{\epsilon_i}{kT_1}\left(\frac{dp_i}{dt}\right)_{\text{int.}}+\frac{\epsilon_i}{kT_2}\left(\frac{dq_i}{dt}\right)_{\text{int.}}\right\}. \tag{28}$$

In the steady state the rate of change of p_i due to internal transitions is just balanced by its rate of change due to molecules passing to or from the second container. Using this fact as expressed in Eq. (26) we obtain the equation

$$\frac{dS}{dt}=Nk\sum_i b_{ii}(q_i^0-p_i^0)\left(\frac{\epsilon_i}{kT_1}-\frac{\epsilon_i}{kT_2}\right), \qquad (29)$$

or

$$\frac{dS}{dt}=N\sum_i b_{ii}(p_i^0-q_i^0)\epsilon_i\left(\frac{1}{T_2}-\frac{1}{T_1}\right). \qquad (30)$$

In this form the entropy production, in the steady state, is expressed as the sum over all states of the product of two factors. The first, $b_{ii}(p_i^0-q_i^0)N$, is the net number of molecules going from the first container to the second per unit time. The second, $(\epsilon_i/T_2)-(\epsilon_i/T_1)$, is the increase in entropy produced when energy ϵ_i is removed from the higher temperature bath (T_1) and delivered to the lower temperature bath (T_2). Hence, the entropy is produced simply by the conduction of heat from the higher to the lower temperature by means of the flow of molecules. The steady state is maintained since the total number of molecules in each container ($N\sum p_i$ and $N\sum q_i$) remain constant in time. [We notice that the expression for dS/dt in Eq. (30) is positive since $p_i^0>q_i^0$ if $T_1>T_2$.]

In conclusion we may say that the theorem of minimum entropy production has been proved, by the methods of statistical mechanics, for a particular type of irreversible process. This proof has helped to clarify the assumptions on which the theorem is based and also its position in the general framework of statistical mechanics.

APPENDIX

This appendix is concerned with the derivation of Eq. (5) of the text which relates the transition probabilities per unit time a_{ij} and a_{ji} for a system in contact with a heat bath at temperature T.

It can be shown by the use of quantum-mechanical perturbation theory that for an isolated system the transition probabilities per unit time are symmetric.[12] More precisely, if we have two nondegenerate states, a and b, each being one of a group of closely spaced states, then the probability per unit time of a transition from a to b is equal to the probability per unit time of a transition from b to a. (The condition of isolation implies that real transitions occur only between states of essentially equal energy.)

We shall use this result to derive Eq. (5) by considering the isolated system consisting of the system of interest plus the heat bath at temperature T with which it is in contact.

We shall denote the states of the system by indices $i, j, \cdots$ and those of the heat bath by indices $\alpha, \beta, \cdots$. If we denote by $W_{i\alpha}$ the probability of finding the combined system in the joint state (i, α), and by $A(i\alpha|j\beta)$ the transition probability per unit time from (i, α) to (j, β), then the equation for $dW_{i\alpha}/dt$ has the form

$$dW_{i\alpha}/dt=\sum_{j,\beta} A(j\beta|i\alpha)W_{j\beta}-W_{i\alpha}\sum_{j,\beta} A(i\alpha|j\beta), \qquad \text{(A-1)}$$

where

$$A(i\alpha|j\beta)=A(j\beta|i\alpha). \qquad \text{(A-2)}$$

The probability w_i of finding the system itself in state i is obtained from the $W_{i\alpha}$ by summing over all heat bath states α; thus we have

$$w_i=\sum_\alpha W_{i\alpha}. \qquad \text{(A-3)}$$

Our problem is to rewrite Eq. (A-1), by summing over α, in such a way as to obtain an equation of the form

$$dw_i/dt=\sum_j a_{ji}w_j-w_i\sum_j a_{ij}, \qquad \text{(A-4)}$$

and to show that

$$a_{ji}e^{-\epsilon_j/kT}=a_{ij}e^{-\epsilon_i/kT}. \qquad \text{(A-5)}$$

In our notation ϵ_i is the energy of state i of the system and E_α is the energy of state α of the heat bath.

Summing on α in Eq. (A-1) leads to the equation

$$dw_i/dt=\sum_{j\beta} W_{j\beta}\sum_\alpha A(j\beta|i\alpha)-\sum_{j\beta}\sum_\alpha A(i\alpha|j\beta)W_{i\alpha}. \qquad \text{(A-6)}$$

We see that Eq. (A-6) is not yet in the desired form of Eq. (A-4). To obtain this form we must assume that

$$W_{i\alpha}=w_i/s_\alpha, \qquad \text{(A-7)}$$

where s_α is the degeneracy of the level α of the heat bath. It is important that we clarify the assumptions contained in Eq. (A-7). The first assumption is that the probability $W_{i\alpha}$ is the product of the probability w_i that the system is in state i and the probability that the heat bath is in state α. The second assumption is that this latter probability is just the reciprocal of the number of bath states of energy E_α. These assumptions are justified by the fact that a heat bath is a system of arbitrarily large heat capacity. Consequently, the probability of finding the bath in a given state is practically unaffected by the coupling to the system.

Using Eq. (A-7) in Eq. (A-6) we obtain the equation

$$\frac{dw_i}{dt}=\sum_j w_j\sum_{\alpha\beta}\frac{1}{s_\beta}A(j\beta|i\alpha)-w_i\sum_j\sum_{\alpha\beta}\frac{1}{s_\alpha}A(i\alpha|j\beta). \qquad \text{(A-8)}$$

Equation (A-8) has the form of Eq. (A-4) with

$$a_{ij}=\sum_{\alpha\beta}\frac{1}{s_\alpha}A(i\alpha|j\beta), \quad a_{ji}=\sum_{\alpha\beta}\frac{1}{s_\beta}A(j\beta|i\alpha). \qquad \text{(A-9)}$$

We must now evaluate s_α and s_β. The density of states in a heat bath varies as the energy raised to an

---

[12] See reference 9, pp. 424–436, and E. C. Kemble, *The Fundamental Principles of Quantum Mechanics* (McGraw-Hill Book Company, Inc., New York, 1937), Chap. XII.

enormous power M, of the order of the number of molecules in the heat bath.[13] Hence, if $\mathcal{E}$ is the energy of the system plus bath we have

$$s_\alpha = cE_\alpha{}^M = c(\mathcal{E}-\epsilon_i)^M \cong c\mathcal{E}^M[1-(M\epsilon_i/\mathcal{E})] \cong c\mathcal{E}^M \exp(-M\epsilon_i/\mathcal{E}), \quad \text{(A-10)}$$

where $\mathcal{E}=\epsilon_i+E_\alpha$. Identifying $\mathcal{E}/M$ with kT we have

$$s_\alpha = c\mathcal{E}^M e^{-\epsilon_i/kT}, \quad s_\beta = c\mathcal{E}^M e^{-\epsilon_j/kT}, \quad \text{(A-11)}$$

since $\mathcal{E}$, the total energy, is also equal to ϵ_j+E_β.

Using Eq. (A-2) with Eq. (A-9) and (A-11), we obtain finally

$$a_{ij}e^{-\epsilon_i/kT} = a_{ji}e^{-\epsilon_j/kT},$$

which is Eq. (A-5).

---

[13] See reference 9, p. 490, and E. Schrödinger, *Statistical Thermodynamics* (Cambridge University Press, Cambridge, 1952), pp. 38 and 89.

PHYSICAL REVIEW VOLUME 96, NUMBER 2 OCTOBER 15, 1954

Flow of Electrons and Holes through the Surface Barrier Region in Point Contact Rectification*

MELVIN CUTLER
Hughes Research and Development Laboratories, Culver City, California

(Received April 29, 1954; revised manuscript received July 8, 1954)

Equations are derived for the flow of electrons and holes through a surface barrier region by using the emission theory. The solution allows for nonequilibrium concentration of carriers on the semiconductor side of the barrier. It also allows for the possibility that part of the applied potential is between the metal surface and the semiconductor surface, as would occur if the surface states do not remain in equilibrium with the metal. The solution for the rectification characteristic is completed for the special case of small currents by combining the barrier region equations with the solution for current flow beyond the barrier region. The resulting equations are compared with those for *p-n* junctions, and the implications are discussed with regard to the relative roles of diffusion and emission in the flow of electrons and holes. Finally, the small current equations are compared with experiment, with a discussion of the evidence for the existence of an intersurface potential.

1. INTRODUCTION

THE early "diode" theory for point contact rectification was based on the flow of a single type of carrier, and did not consider the effects of minority carrier injection.[1,2] After the discovery of the latter phenomenon, the small current theory for *p-n* junctions was advanced by Shockley.[3] The application of this theory to point contacts was also discussed.[4]

The present theory of point contact rectifiers does not agree with experiment in the current voltage characteristic. The discrepancies have been discussed extensively.[5,6] particularly with reference to the reverse characteristic.[7] In view of this disagreement, it seems desirable to develop the theory in a more systematic and rigorous fashion. In such a manner, one may be able to investigate more carefully the validity of the accepted model for surface rectification.

The behavior of a rectifier may be obtained from the combination of the solutions of two separate problems: The flow of electrons and holes in the barrier region, which is "emission" controlled; and the flow beyond the barrier region, which is diffusion controlled.[8] A solution of the latter kind, which is one dimensional and applicable to *p-n* junctions or large area surface contacts has been derived by Van Roosbroeck.[9] The purpose of this paper is to present a self-contained solution to the former problem; that is, for the emission of electrons and holes through the barrier region of a semiconductor. In addition the solution will be completed for small currents in point contacts, for which case the diffusion problem has been solved; the results are compared with *p-n* junction theory. A general solution for the diffusion problem with radial flow, in combination with the emission equations, would permit the discussion of the forward (large current) characteristic of point contact rectifiers. This will be done in a later paper.

The emission equations to be derived are more general than the original diode equations in two ways.

---

* This work was first presented at the Cambridge American Physical Society Meeting in February, 1953 [Phys. Rev. 90, 337 (1954)].

[1] R. C. Torrey and C. A. Whitmer, *Crystal Rectifiers* (McGraw-Hill Book Company, Inc., New York, 1948), Chap. 4.

[2] N. F. Mott and R. W. Gurney, *Electronic Processes in Ionic Crystals* (Oxford University Press, London, 1948), Chap. 5.

[3] W. Shockley, *Electrons and Holes in Semiconductors* (D. Van Nostrand Company, Inc., New York, 1950), Chap. 12.

[4] See reference 3, pp. 95–101.

[5] S. Benzer, J. Appl. Phys. **20**, 804 (1949).

[6] J. Bardeen and W. H. Brattain, Phys. Rev. **75**, 1208 (1949).

[7] J. H. Simpson and H. L. Armstrong, J. Appl. Phys. **24**, 25 (1953).

[8] This approach is discussed in a paper which has appeared recently. See P. C. Banbury, Proc. Phys. Soc. (London) **B66**, 833 (1953).

[9] W. Van Roosbroeck, Bell System Tech. J. **29**, 560 (1950).