

This suggests that  $f_j$  in (3.26) be interpreted as the information content of the symbol with index  $j$  and  $p_j$  as the probability (or relative frequency).

Now let us assume that a set of measurements has led us to the relative frequency  $p_j$  for the symbol with index  $j$ . Then let us assume that – possibly under different conditions – we determine a new relative frequency  $p'_j$ . What is the corresponding change of information,  $\Delta_j$ ?

Adopting the interpretation of (3.26), we are immediately led to define it by

$$\Delta_j = K \ln p'_j - K \ln p_j . \quad (3.27)$$

To obtain the mean change of information, we average (3.27) over the new distribution function (“relative frequency”)  $p'_j$ . We thus obtain the so-called information gain (or “Kullback information”).

$$K(p', p) = \sum_j p'_j \Delta_j = K \sum_j p'_j \ln \frac{p'_j}{p_j} , \quad (3.28)$$

where, of course,

$$\sum_j p_j = 1 \quad \text{and} \quad (3.29)$$

$$\sum_j p'_j = 1 . \quad (3.30)$$

The information gain  $K(p', p)$  has the following important property.

$$K(p', p) \geq 0 . \quad (3.31)$$

The equality sign holds if and only if

$$p' \equiv p, \text{ i.e. } p'_k = p_k \quad \text{for all } k\text{'s} .$$

### 3.3 Information Entropy and Constraints

In this section and in the next chapter we will be concerned, in particular, with applications of the information concept to physics and shall thus follow the convention of denoting the information by  $S$ , and identifying the constant  $K$  in (3.3) with Boltzmann's constant  $k_B$ . For reasons which will appear later,  $S$  will be called information entropy. Because chemical and biological systems can be viewed as physical systems, our considerations apply equally well to these systems too. The general formalism of this chapter is also applicable to other sciences, such as information processing, etc. We start from the basic expression

$$S = -k_B \sum_i p_i \ln p_i . \quad (3.32)$$

The indices  $i$  may be considered as describing individual features of the particles or subsystems. Let us explain this in some detail. The index  $i$  may describe, for



instance, the position of a gas particle or it may describe its velocity or both properties. In our previous examples the index  $i$  referred to boxes filled with balls. In a more general interpretation the index  $i$  represents the values that a random variable may acquire. In this section we assume for simplicity that the index  $i$  is discrete.

A central task to be solved in this book consists in finding ways to determine the  $p_i$ 's (compare for example the gas molecules in a container where one wants to know their location). The problem we are confronted with in many disciplines is to make unbiased estimates leading to  $p_i$ 's which are in agreement with all the possible knowledge available about the system. Consider an ideal gas in one dimension. What we could measure, for instance, is the center of gravity. In this case we would have as constraint an expression of the form

$$\sum_i p_i q_i = M \quad (3.33)$$

where  $q_i$  measures the position of the cell  $i$ .  $M$  is a fixed quantity equal to  $Q/N$ , where  $Q$  is the coordinate of the center of gravity, and  $N$  the particle number. There are, of course, very many sets of  $p_i$ 's, which fulfill the relation (3.33). Thus we could choose a set  $\{p_i\}$  rather arbitrarily, i.e., we would favor one set against another one. Similar to ordinary life, this is a biased action. How may it be unbiased? When we look again at the example of the gas atoms, then we can invoke the principle stated in Sect. 3.1. With an overwhelming probability we will find those distributions realized for which (3.32) is a maximum. However, due to (3.33) not all distributions can be taken into account. Instead we have to seek the maximum of (3.32) under the constraint (3.33). This principle can be generalized if we have a set of constraints. Let, for example, the variable  $i$  distinguish between different velocities. Then we may have the constraint that the total kinetic energy  $E_{\text{kin}}^{\text{tot}}$  of the particles is fixed. Denoting the kinetic energy of a particle with mass  $m$  and velocity  $v_i$  by  $f_i$  [ $f_i = (m/2)v_i^2$ ] the mean kinetic energy per particle is given by

$$\sum_i p_i f_i = E_{\text{kin}} \quad (3.33a)$$

In general the single system  $i$  may be characterized by quantities  $f_i^{(k)}$ ,  $k = 1, 2, \dots, M$  (position, kinetic energy or other typical features). If these features are additive, and the corresponding sums are kept fixed at values  $f_k$ , then the constraints take the form

$$\sum_i p_i f_i^{(k)} = f_k \quad (3.34)$$

We further add the usual constraint that the probability distribution is normalized

$$\sum_i p_i = 1 \quad (3.35)$$

The problem of finding the extremum of (3.32) under the constraints (3.34) and (3.35) can be solved by using the method of Lagrange multipliers  $\lambda_k$ ,  $k = 1, 2, \dots, M$  (cf. Sect. 3.1). We multiply the left-hand side of (3.34) by  $\lambda_k$  and the left-hand side of

(3.35) by  $(\lambda - 1)$  and take the sum of the resulting expressions. We then subtract this sum from  $(1/k_B)S$ . The factor  $1/k_B$  amounts to a certain normalization of  $\lambda, \lambda_k$ . We then have to vary the total sum with respect to the  $p_i$ 's

$$\delta \left[ \frac{1}{k_B} S - (\lambda - 1) \sum_i p_i - \sum_k \lambda_k \sum_i p_i f_i^{(k)} \right] = 0 . \quad (3.36)$$

Differentiating with respect to  $p_i$  and setting the resulting expression equal to zero, we obtain

$$-\ln p_i - 1 - (\lambda - 1) - \sum_k \lambda_k f_i^{(k)} = 0 , \quad (3.37)$$

which can be readily solved for  $p_i$  yielding

$$p_i = \exp \left( -\lambda - \sum_k \lambda_k f_i^{(k)} \right) . \quad (3.38)$$

Inserting (3.38) into (3.35) yields

$$e^{-\lambda} \sum_i \exp \left( -\sum_k \lambda_k f_i^{(k)} \right) = 1 . \quad (3.39)$$

It is now convenient to abbreviate the sum over  $i$ ,  $\sum_i$  in (3.39) by

$$\sum_i \exp \left( -\sum_k \lambda_k f_i^{(k)} \right) = Z(\lambda_1, \dots, \lambda_M) , \quad (3.40)$$

which we shall call the partition function. Inserting (3.40) into (3.39) yields

$$e^{-\lambda} = Z \quad \text{or} \quad (3.41)$$

$$\lambda = \ln Z , \quad (3.42)$$

which allows us to determine  $\lambda$  once the  $\lambda_k$ 's are determined. To find equations for the  $\lambda_k$ 's we insert (3.38) into the equations of the constraints (3.34) which lead immediately to

$$\langle f_i^{(k)} \rangle = \sum_i p_i f_i^{(k)} = e^{-\lambda} \sum_i \exp \left( -\sum_l \lambda_l f_i^{(l)} \right) f_i^{(k)} . \quad (3.43)$$

Equation (3.43) has a rather similar structure to (3.40). The difference between these two expressions arises because in (3.43) each exponential function is still multiplied by  $f_i^{(k)}$ . However, we may easily derive the sum occurring in (3.43) from (3.40) by differentiating (3.40) with respect to  $\lambda_k$ . Expressing the first factor in (3.43) by  $Z$  according to (3.41) we thus obtain

$$\langle f_i^{(k)} \rangle = \frac{1}{Z} \left( -\frac{\partial}{\partial \lambda_k} \right) \underbrace{\sum_i \exp \left( -\sum_l \lambda_l f_i^{(l)} \right)}_Z \quad (3.44)$$



or in still shorter form

$$f_k \equiv \langle f_i^{(k)} \rangle = -\frac{\partial \ln Z}{\partial \lambda_k} . \quad (3.45)$$

Because the  $f_k$  on the left-hand side are prescribed [compare (3.34)] and  $Z$  is given by (3.40) which is a function of the  $\lambda_k$ 's in a special form, (3.45) is a concise form for a set of equations for the  $\lambda_k$ 's.

We further quote a formula which will become useful later on. Inserting (3.38) into (3.32) yields

$$\frac{1}{k_B} S_{\max} = \lambda \sum_i p_i + \sum_k \lambda_k \sum_i p_i f_i^{(k)} \quad (3.46)$$

which can be written using (3.34) and (3.35) as

$$\boxed{\frac{1}{k_B} S_{\max} = \lambda + \sum_k \lambda_k f_k} . \quad (3.47)$$

The maximum of the information entropy may thus be represented by the mean values  $f_k$  and the Lagrange multipliers  $\lambda_k$ . Those readers who are acquainted with the Lagrange equations of the first kind in mechanics will remember that the Lagrange multipliers have a physical meaning, in that case, of forces. In a similar way we shall see later on that the Lagrange multipliers  $\lambda_k$  have physical (or chemical or biological, etc) interpretations. In deriving the above formulas [i.e., (3.38,42) with (3.32,45,47)] we have completed our original task of finding the  $p$ 's and  $S_{\max}$ .

We now derive some further useful relations. We first investigate how the information  $S_{\max}$  is changed if we change the functions  $f_i^{(k)}$  and  $f_k$  in (3.34). Because  $S$  depends, according to (3.47), not only on the  $f$ 's but also on  $\lambda$  and the  $\lambda_k$ 's which are functions of the  $f$ 's, we must exercise some care in taking the derivatives with respect to the  $f$ 's. We therefore first calculate the change of  $\lambda$  (3.42)

$$\delta \lambda \equiv \delta \ln Z = \frac{1}{Z} \delta Z .$$

Inserting (3.40) for  $Z$  yields

$$\delta \lambda = e^{-\lambda} \sum_i \sum_k \{ -\delta \lambda_k f_i^{(k)} - \lambda_k \delta f_i^{(k)} \} \exp \left( -\sum_l \lambda_l f_i^{(l)} \right)$$

which, from the definition of  $p_i$  (3.38) transforms to

$$\delta \lambda = -\sum_k \left[ \delta \lambda_k \sum_i p_i f_i^{(k)} + \lambda_k \sum_i p_i \delta f_i^{(k)} \right] .$$

Equation (3.43) and an analogous definition of  $\langle \delta f_i^{(k)} \rangle$  allow us to write the last line as

$$-\sum_k [\delta \lambda_k \langle f_i^{(k)} \rangle + \lambda_k \langle \delta f_i^{(k)} \rangle] . \quad (3.48)$$

Inserting this into  $\delta S_{\max}$  from (3.47), we find that the variation of the  $\lambda_k$ 's drops out and we are left with

$$\delta S_{\max} = k_B \sum_k \lambda_k [\delta \langle f_i^{(k)} \rangle - \langle \delta f_i^{(k)} \rangle] . \quad (3.49)$$

We write this in the form

$$\delta S_{\max} = k_B \sum_k \lambda_k \delta Q_k \quad (3.50)$$

where we define a "generalized heat" by means of

$$\delta Q_k = \delta \langle f_i^{(k)} \rangle - \langle \delta f_i^{(k)} \rangle . \quad (3.51)$$

The notation "generalized heat" will become clearer below when contact with thermodynamics is made. In analogy to (3.45), a simple expression for the variance of  $f_i^{(k)}$  may be derived:

$$\langle f_i^{(k)2} \rangle - \langle f_i^{(k)} \rangle^2 = \frac{\partial^2 \ln Z}{\partial \lambda_k^2} . \quad (3.52)$$

In many practical applications,  $f_i^{(k)}$  depends on a further quantity  $\alpha$  (on a set of such quantities  $\alpha_1, \alpha_2, \dots$ ). Then we want to express the change of the mean value (3.34), when  $\alpha$  is changed. Taking the derivative of  $f_{i,\alpha}^{(k)}$  with respect to  $\alpha$  and taking the average value, we find

$$\left\langle \frac{\partial f_{i,\alpha}^{(k)}}{\partial \alpha} \right\rangle = \sum_i p_i \frac{\partial f_{i,\alpha}^{(k)}}{\partial \alpha} . \quad (3.53)$$

Using the  $p_i$ 's in the form (3.38) and using (3.41), the right-hand side of (3.53) may be written in the form

$$\frac{1}{Z} \sum_i \frac{\partial f_{i,\alpha}^{(k)}}{\partial \alpha} \exp \left( - \sum_j \lambda_j f_{i,\alpha}^{(j)} \right) , \quad (3.54)$$

which may easily be expressed as a derivative of  $Z$  with respect to  $\alpha$ :

$$(3.54) = - \frac{1}{Z} \frac{1}{\lambda_k} \frac{\partial Z}{\partial \alpha} . \quad (3.55)$$

Thus we are lead to the final formula

$$\boxed{ - \frac{1}{\lambda_k} \frac{\partial \ln Z}{\partial \alpha} = \left\langle \frac{\partial f_{i,\alpha}^{(k)}}{\partial \alpha} \right\rangle } . \quad (3.56)$$

If there are several parameters  $\alpha_i$  present, this formula can be readily generalized by writing  $\alpha_i$  in place of  $\alpha$  in (3.56).

As we have seen several times, the quantity  $Z$ , (3.40), or its logarithm, is very useful [see e.g., (3.45,52,56). We want to convince ourselves that  $\ln Z \equiv \lambda$  [cf. (3.42)]



may be directly determined by a variational principle. A glance at (3.36) reveals that (3.36) can also be interpreted in the following way: Seek the extremum of

$$\frac{1}{k_B} S - \sum_k \lambda_k \sum_i p_i f_i^{(k)} \quad (3.57)$$

under the only constraint

$$\sum p_i = 1 . \quad (3.58)$$

Now, by virtue of (3.34,47,42) the extremum of (3.57) is indeed identical with  $\ln Z$ . Note that the spirit of the variational principle for  $\ln Z$  is different from that for  $S$ . In the former case, we had to seek the maximum of  $S$  under the constraints (3.34,35) with  $f_k$  fixed and  $\lambda_k$  unknown. Here, only one constraint, (3.58), applies and the  $\lambda_k$ 's are assumed as given quantities. How such a switching from one set of fixed quantities to another one can be done will become more evident by the example from physics given in Chap. 4, which will also elucidate many other aspects of the foregoing discussion.

### 3.4 Continuous Variables

In most applications that we have in mind, the variables  $\xi$  are not discrete but continuous. One may then easily convince oneself that in such a case, at least in general, the information diverges. This is due to the fact that we have continuously many states. Therefore we have to discuss briefly how we can define information for continuous variables. We start from the definition of the probability density given by

$$\text{Prob}(\chi \leq \xi \leq \chi + \Delta\xi) = P(\xi) \Delta\xi . \quad (3.59)$$

We now invoke the idea that measurements can be made only with finite accuracy. Therefore we introduce an interval of accuracy and define a new probability distribution by

$$P_\varepsilon(j) = \int_{\xi_j - \varepsilon/2}^{\xi_j + \varepsilon/2} P(\xi) d\xi . \quad (3.60)$$

Assuming that  $P(\xi)$  is continuous we may approximate (3.60) by

$$P_\varepsilon(j) \approx P(\xi_j) \varepsilon . \quad (3.61)$$

We define the information with respect to the interval of accuracy  $\varepsilon$  by means of

$$I_\varepsilon = - \sum_j P_\varepsilon(j) \ln P_\varepsilon(j) . \quad (3.62)$$

Inserting (3.61) into (3.62) we obtain

$$I_{\varepsilon} = - \sum_j \varepsilon P(\xi_j) \ln P(\xi_j) - \underbrace{\sum_j \varepsilon P(\xi_j)}_{=1} \ln \varepsilon \quad (3.63)$$

so that our final result reads

$$I_{\varepsilon} = - \int d\xi P(\xi) \ln P(\xi) - \ln \varepsilon . \quad (3.64)$$

In the following we shall drop the constant and uninteresting term  $-\ln \varepsilon$ . The extension to several variables is obvious.

## 4. An Example from Physics: Thermodynamics

To visualize the meaning of the index  $i$ , let us identify it with the velocity of a particle. In a more advanced theory  $p_i$  is the occupation probability of a quantum state  $i$  of a many-particle system. Further, we identify  $f_{i,\alpha}^{(k)}$  with energy  $E$  and the parameter  $\alpha$  with the volume. Thus we put

$$f_{i,\alpha}^{(k)} = E_i(V) ; \quad k = 1 , \quad (4.1)$$

and have the identifications

$$f_1 \leftrightarrow U \equiv \langle E_i \rangle ; \quad \alpha \leftrightarrow V ; \quad \lambda_1 = \beta . \quad (4.2)$$

We have, in particular, set  $\lambda_1 = \beta$ . With this, we may write a number of the previous formulas in a way which can be immediately identified with relations well known in thermodynamics and statistical mechanics. Instead of (3.38) we find

$$p_i = \exp[-\lambda - \beta E_i(V)] \quad (4.3)$$

which is the famous Boltzmann distribution function.

Equation (3.47) acquires the form

$$\frac{1}{k_B} S_{\max} = \ln Z + \beta U \quad (4.4)$$

or, after a slight rearrangement of this equation

$$U - \frac{1}{k_B \beta} S_{\max} = -\frac{1}{\beta} \ln Z . \quad (4.5)$$

This equation is well known in thermodynamics and statistical physics. The first term may be interpreted as the internal energy  $U$ ,  $1/\beta$  as the absolute temperature  $T$  multiplied by Boltzmann's constant  $k_B$ .  $S_{\max}$  is the entropy. The right-hand side represents the free energy,  $\mathcal{F}$ , so that in thermodynamic notation (4.5) reads

$$U - TS = \mathcal{F} . \quad (4.6)$$

By comparison we find

$$\mathcal{F} = -k_B T \ln Z , \quad (4.7)$$

and  $S = S_{\max}$ . Therefore we will henceforth drop the suffix "max". Equation (3.40)



now reads

$$Z = \sum_i e^{-\beta E_i} \quad (4.8)$$

and is nothing but the usual partition function. A number of further identities of thermodynamics can easily be checked by applying the above formulas.

The only problem requiring some thought is the identification of independent and dependent variables. Let us begin with the information entropy,  $S_{\max}$ . In (3.47) it appears as a function of  $\lambda$ ,  $\lambda_\mu$  and the  $f_k$ . However,  $\lambda$  and the  $\lambda_\mu$  are themselves determined by equations which contain the  $f_k$  and  $f_i^{(k)}$  as given quantities [cf. (3.40,42,43)]. Therefore, the independent variables are  $f_k$  and  $f_i^{(k)}$ , and the dependent variables are  $\lambda$  and  $\lambda_k$ , and thus, by virtue of (3.47),  $S_{\max}$ . In practice the  $f_i^{(k)}$  are fixed functions of  $i$  (e.g., the energy of state “ $i$ ”), but still depending on parameters  $\alpha$  [e.g., the volume, cf. (4.1)]. Thus the truly independent variables in our approach are the  $f_k$  (as above) and the  $\alpha$ 's. In conclusion we thus find  $S = S(f_k, \alpha)$ . In our example,  $f_1 = E \equiv U$ ,  $\alpha = V$ , and therefore

$$S = S(U, V) . \quad (4.9)$$

Now let us apply the general relation (3.49) to our specific model. If we vary only the internal energy,  $U$ , but leave  $V$  unchanged, then

$$\delta \langle f_i^{(1)} \rangle \equiv \delta f_1 \equiv \delta U \neq 0 , \quad \text{and} \quad (4.10)$$

$$\delta f_{i,\alpha}^{(1)} \equiv \delta E_i(V) = \frac{\delta E_i(V)}{\delta V} \delta V = 0 , \quad \text{and therefore} \quad (4.11)$$

$$\delta S = k_B \lambda_1 \delta U \quad \text{or}$$

$$\frac{\delta S}{\delta U} = k_B \lambda_1 (\equiv k_B \beta) . \quad (4.12)$$

According to thermodynamics, the left-hand side of (4.12) defines the inverse of the absolute temperature

$$\frac{\delta S}{\delta U} = \frac{1}{T} . \quad (4.13)$$

This yields  $\beta = 1/(k_B T)$  as anticipated above. On the other hand, varying  $V$  but leaving  $U$  fixed, i.e.,

$$\delta \langle f_i^{(1)} \rangle = 0 , \quad \text{but} \quad (4.14)$$

$$\langle \delta f_i^{(1)} \rangle = \left\langle \frac{\delta E_i(V)}{\delta V} \right\rangle \delta V \neq 0 \quad (4.15)$$

yields in (3.49)

$$\delta S = k_B(-\lambda_1) \left\langle \frac{\delta E_i(V)}{\delta V} \right\rangle \delta V \quad \text{or}$$

$$\frac{\delta S}{\delta V} = -\frac{1}{T} \left\langle \frac{\delta E_i(V)}{\delta V} \right\rangle . \quad (4.16)$$

Since thermodynamics teaches us that

$$\frac{\delta S}{\delta V} = \frac{P}{T} \quad (4.17)$$

where  $P$  is the pressure, we obtain by comparison with (4.16)

$$\left\langle \frac{\delta E_i(V)}{\delta V} \right\rangle = -P . \quad (4.18)$$

Inserting (4.13,17) into (3.49) yields

$$\delta S = \frac{1}{T} \delta U + \frac{1}{T} P \delta V . \quad (4.19)$$

In thermodynamics the right-hand side is equal to  $dQ/T$  where  $dQ$  is heat. This explains the notation “generalized heat” used after (3.51). These considerations may be generalized to different kinds of particles whose average numbers  $N_k$ ;  $k = 1, \dots, m$ , are prescribed quantities. We therefore identify  $f_1$  with  $E$ , but  $f_{k'+1}$  with  $N_{k'}$ ;  $k' = 1, \dots, m$  (note the shift of index!). Since each kind of particle,  $l$ , may be present with different numbers  $N_l$  we replace the index  $i$  by  $i, N_1, \dots, N_m$  and put

$$f_i^{(k+1)} \rightarrow f_{i, N_1, \dots, N_m}^{(k+1)} = N_k .$$

To be in accordance with thermodynamics, we put

$$\lambda_{k+1} = -\frac{1}{k_B T} \mu_k , \quad (4.20)$$

where  $\mu_k$  is called the chemical potential.

Equation (3.47) with (3.42) acquires (after multiplying both sides by  $k_B T$ ) the form

$$TS = \underbrace{k_B T \ln Z}_{-\mathcal{F}} + U - \mu_1 \bar{N}_1 - \mu_2 \bar{N}_2 - \dots - \mu_m \bar{N}_m . \quad (4.21)$$

Equation (3.49) permits us to identify

$$\frac{\delta S}{\delta \bar{N}_k} = -k_B \lambda_{k+1} = \frac{1}{T} \mu_k . \quad (4.22)$$



The partition function reads

$$Z = \sum_{N_1 N_2 \dots N_m} \sum_i \exp \left\{ -\frac{1}{k_B T} [E_i(V) - \mu_1 N_1 - \dots - \mu_m N_m] \right\}. \quad (4.23)$$

While the above considerations are most useful for irreversible thermodynamics, in thermodynamics the role played by independent and dependent variables is, to some extent, exchanged. It is not our task to treat these transformations which give rise to the different thermodynamic potentials (Gibbs, Helmholtz, etc). We just mention one important case: Instead of  $U$ ,  $V$  (and  $N_1, \dots, N_m$ ), as independent variables, one may introduce  $V$  and  $T = (\partial S / \partial U)^{-1}$  (and  $N_1, \dots, N_m$ ) as new independent variables. As an example we treat the  $U$ - $V$  case (putting formally  $\mu_1, \mu_2, \dots = 0$ ). According to (4.7) the free energy,  $\mathcal{F}$ , is there directly given as a function of  $T$ . The differentiation  $\partial \mathcal{F} / \partial T$  yields

$$-\frac{\partial \mathcal{F}}{\partial T} = k_B \ln Z + \frac{1}{T} \frac{1}{Z} \sum_i E_i e^{-\beta E_i}.$$

The second term on the right-hand side is just  $U$ , so that

$$-\frac{\partial \mathcal{F}}{\partial T} = k_B \ln Z + \frac{1}{T} U. \quad (4.24)$$

Comparing this relation with (4.5), where  $1/\beta = k_B T$ , yields the important relation

$$-\frac{\partial \mathcal{F}}{\partial T} = S \quad (4.25)$$

where we have dropped the suffix "max".

Readers who are interested in the application of the above formalism to irreversible thermodynamics, i.e. to *relaxation phenomena*, are referred to my book *Synergetics. An Introduction*. In the present book we shall be concerned with quite a different field, namely nonequilibrium phase transitions in physical and biological systems far from thermal equilibrium and in nonphysical systems (e.g. economy).

## **5. Application of the Maximum Information Principle to Self-Organizing Systems**

### **5.1 Introduction**

According to Chap.1, self-organizing systems are systems which can acquire macroscopic spatial, temporal, or spatio-temporal structures by means of internal processes without specific interference from the outside. Hitherto, the distribution functions of the order parameters governing the macroscopic structures could only be calculated by microscopic theories (cf. Chap.2). In the present section we derive them from macroscopic quantities, and we demonstrate this procedure explicitly by means of the single and multimode laser close to the lasing threshold.

The maximum information entropy principle allows one to make unbiased estimates on the probability distribution of microscopic states of systems of which otherwise only certain averages, corresponding to macroscopic observations, are known. As we have seen in the preceding section this principle provides one with a very elegant access to many of the basic relations and concepts of thermodynamics, i.e. it can be applied very nicely to systems in thermal equilibrium. On the other hand no successful attempts are known of a general application of this principle to systems far from thermal equilibrium.

In this section I wish to show how the maximum information entropy principle can indeed be very successfully applied to nonequilibrium systems provided they acquire macroscopic structures through self-organization. In this way we shall recover well-known distribution functions of such systems. These functions have been previously derived from microscopic theories (cf. Sects. 2.4,7). At the same time generalizations of these functions now become available, too. In order to illustrate our procedure we first focus our attention on lasers.

### **5.2 Application to Self-Organizing Systems: Single Mode Laser**

Self-organizing systems are characterized by the occurrence of macroscopic structures which can be described by adequate order parameters. Instead of introducing abstract considerations we prefer to illustrate our procedure by means of explicit examples. It will be important to strictly stick to the basic notion of the maximum information entropy principle, namely to consider the macroscopically observed quantities.

The quantities observed experimentally of a single mode laser are its intensity and the second moment of the intensity in the steady state case. It is well known



from laser theory that the basic difference between the light from lamps and that from lasers becomes apparent only when the second moments of the intensity are measured in addition to the first moment. Further measured quantities are intensity correlations, but because we have a time-independent theory in mind we shall ignore this information here.

The space- and time-dependent electric field strength of a single mode laser can be written in the form

$$E(x, t) = E(t) \sin kx , \quad (5.1)$$

where the amplitude  $E(t)$  can be decomposed into its positive and negative frequency part according to

$$E(t) = \underbrace{B e^{-i\omega t}}_b + \underbrace{B^* e^{i\omega t}}_{b^*} . \quad (5.2)$$

If we measure the intensity of the light field over time intervals large compared to an oscillation period, but small compared to the fluctuation times of  $B(t)$ , the output intensity is proportional to  $B^*B$  and to the loss rate,  $2\kappa$ , of the laser. For the sake of simplicity we drop all other constants and put

$$I = 2\kappa B^*B . \quad (5.3)$$

Similarly, the intensity squared, if averaged over the same time interval, turns out to be

$$I^2 = 4\kappa^2 B^{*2} B^2 . \quad (5.4)$$

Because of the fluctuations of the laser,  $B^*$  and  $B$  are random variables which belong to a stationary process. This leads us to identify  $B, B^*$  with the indices  $i$  of  $p_i$  in (3.32), where we put  $k_B = 1$  and interpret the right-hand side as information,  $i$ . Because the random variables  $B$  are no longer discrete but continuous we must replace the summation over the indices  $i$  by an integration

$$i = - \int p(B, B^*) \ln p(B, B^*) d^2 B . \quad (5.5)$$

Equation (3.34) may be interpreted as integrals over  $d^2 B$  with the probability  $p(B, B^*)$  as weight functions. Denoting these averages by brackets we are led to consider the following two constraints

$$f_1 = \langle 2\kappa B^*B \rangle , \quad (5.6)$$

$$f_2 = \langle 4\kappa^2 B^{*2} B^2 \rangle . \quad (5.7)$$

Furthermore by the same analogy we are led to define  $f_i^{(k)}$  by means of

$$f_{B, B^*}^{(1)} = 2\kappa B^*B , \quad (5.8)$$

$$f_{B, B^*}^{(2)} = 4\kappa^2 B^{*2} B^2 . \quad (5.9)$$

We are now in a position to apply formula (3.38) immediately and find

$$p(B, B^*) = \exp[-\lambda - \lambda_1 2\kappa B^* B - \lambda_2 4\kappa^2 (B^* B)^2] , \quad (5.10)$$

which in a somewhat different notation reads

$$p(B, B^*) = N \exp(-\alpha |B|^2 - \beta |B|^4) . \quad (5.11)$$

This function is well known in laser physics. It was derived by Risken by solving the Fokker-Planck equation belonging to the laser Langevin equation derived previously by the present author.

We note that in the laser case  $\alpha$  must be negative. But close to threshold  $\alpha$  can take both negative and positive values.

### 5.3 Multimode Laser Without Phase Relations

In this case the field strength is decomposed into its modes according to

$$E(x, t) = \sum_l E_l(t) \sin k_l x , \quad (5.12)$$

where for simplicity we consider only axial modes. Again the mode amplitudes can be decomposed into their positive and negative frequency parts

$$E_l(t) = B_l(t)e^{-i\omega_l t} + B_l^*(t)e^{i\omega_l t} . \quad (5.13)$$

The intensity averaged over time intervals long compared to an oscillation period and short to fluctuation periods is given by

$$I_l \sim 2\kappa_l B_l^* B_l \sim n_l . \quad (5.14)$$

The extension of the results of Sect. 5.2 is straightforward provided we now consider either  $n_l$  or equivalently  $B_l^*$  and  $B_l$  as stochastic variables. We obtain

$$f_l = \langle n_l \rangle , \quad (5.15)$$

$$f_{l,l'} = \langle n_l, n_{l'} \rangle , \quad (5.16)$$

and identifying  $k$  and  $i$  according to

$$k \leftrightarrow \begin{cases} l \\ l, l' \end{cases}$$

$$i \leftrightarrow (n_1, n_2, \dots, n_M) = \mathbf{n} \quad (5.17)$$

we have

$$f_n^{(l)} = n_l , \quad (5.18)$$



$$f_n^{(l,l')} = n_l n_{l'} , \quad (5.19)$$

$$p_i \rightarrow p(n) . \quad (5.20)$$

The application of (3.38) is now straightforward and yields

$$p(n) \equiv p(n_1, \dots, n_M) = \exp \left( -\lambda - \sum_l \lambda_l n_l - \sum_{ll'} \lambda_{ll'} n_l n_{l'} \right) \quad (5.21)$$

as final result. Equivalently (5.21) can be written as

$$N \exp \left( -\sum_l \alpha_l n_l - \sum_{ll'} \beta_{ll'} n_l n_{l'} \right) . \quad (5.22)$$

This form can be derived from a multimode Fokker-Planck equation in special cases in which the solution can be explicitly constructed by means of the principle of detailed balance.

## 5.4 Processes Periodic in Order Parameters

We wish to show how guesses on distribution functions can be made if the processes considered are periodic in the order parameters.

Let us consider as a specific example the angle coordinate  $\phi$  and let us consider moments which are periodic with  $2\pi$ . Then it would seem sensible, instead of the moments of  $\phi$ , to consider the corresponding moments of periodic functions i.e.

$$\langle \sin \phi \rangle , \quad \langle \cos \phi \rangle , \dots \quad (5.23)$$

or, more generally,

$$\langle \sin n\phi \rangle , \quad \langle \cos n\phi \rangle , \quad (5.24)$$

where  $n$  is a positive integer.

In order to illustrate our procedure let us consider the special case in which for symmetry reasons

$$\langle \sin n\phi \rangle = 0 , \quad \text{for all } n . \quad (5.25)$$

In our treatment above we retained only the first few moments. If, in analogy to that, we keep only the first two terms of (5.24), we readily obtain

$$P(\phi) = \exp(\lambda + \lambda_1 \cos \phi + \lambda_2 \cos 2\phi) = \exp[V(\phi)] . \quad (5.26)$$

As we have shown explicitly, the maximum information entropy principle allows us to derive the general form of distribution functions of a nonequilibrium system, such as the laser, in a quite straightforward way. The results agree with distribution functions obtained from microscopic theories under certain restricting conditions.

It is now fairly obvious how one should proceed in other cases. The total state of the system  $q(x, t)$  must be projected onto functions which describe the observed macroscopic spatial or temporal pattern. In this way amplitudes are obtained of which moments can be measured and thus may serve as the functions defining  $f_k$  as well as  $f_i^{(k)}$ .

In spite of the success of our application of the maximum information entropy principle we must bear in mind that only little can be said about the Lagrange multipliers which can, of course, now be determined experimentally. On the other hand it has been the advantage of the microscopic theory that these constants can be determined from first principles and therefore in particular it could be predicted that  $\alpha_i$  changes sign at instability points. However, it may also become possible to deduce such properties from a macroscopic theory.

Our approach can be applied to a number of problems such as convection in fluids, pattern formation in chemical reactions, and growth of morphogenetic fields.

In spite of the formal resemblance of our results to some of thermodynamics, there are still basic differences. First of all we realize that the constants have a physical meaning very different from those for systems in thermal equilibrium. For instance in nonequilibrium systems, such as lasers, we have to deal with output intensities whereas in equilibrium systems we deal with e.g. energies. This is also clearly reflected by a treatment on the microscopic level. While in the microscopic treatment of equilibrium systems energies play a decisive role, in nonequilibrium systems rate constants and growth rates determine the evolving patterns.

We may draw a number of rather far reaching conclusions. Until now, the maximum information (entropy) principle had been applied to thermodynamics and irreversible thermodynamics, but not to nonequilibrium phase transitions. Among the constraints used in the former two fields is the energy. In the present case we deal with the *output intensity*  $I$ . But what is still more important, we now have to include the intensity correlation in the form of the second moment, i.e.  $\langle I^2 \rangle$ . This is never done in equilibrium thermodynamics in the context of the maximum entropy principle. But now we see that the inclusion of  $\langle I^2 \rangle$  is quite obvious and necessary for nonequilibrium phase transitions. At or above threshold,  $\lambda_1 \geq 0$ , and the integral over  $\exp(\lambda_1 |b|^2)$  will diverge, reflecting the effect of critical fluctuations. Because of these, and in order to take care of their limitation due to saturation,  $\langle I^2 \rangle$  must be taken into account. In other words, close to nonequilibrium phase transitions fluctuations become "observables" and must be taken into account by adequate constraints. It appears to me a safe bet that the same is true for phase transitions of systems in thermal equilibrium and that an extension of the maximum information entropy principle is also required there.



## 14. Concluding Remarks and Outlook

In this concluding chapter I wish to discuss what we have achieved in this book and to point out some areas where open questions remain.

In Chap. 1 we dealt with the nature of complex systems. Their most salient feature seems to be that they are practically inexhaustible with respect to our efforts to fully understand them. We must be content to study specific aspects of these systems which seem to us particularly suited for our purposes or interests. Accordingly, we have made an attempt in this book to cope with complex systems by choosing a rather general feature: we have focussed our attention on those situations in which complex systems change their macroscopic behavior qualitatively. More precisely, we have studied situations in which a so-called nonequilibrium phase transition occurs. This approach does not only allow us to deal with physical systems such as lasers and fluids but also with far more complex systems, such as biological systems. Undoubtedly a great variety of further applications can be found. As is witnessed by the laser example, our approach also enables us to treat oscillatory phenomena, but the extension to chaotic systems remains an open question. It is certainly a field for further research. The vehicle we have used in our approach is information. As we have seen, the word information may have quite different meanings so that we first had to discuss various definitions of the word. In particular we saw that the concept of *information* in the Shannon sense seems to be more appropriate in some cases than that of *entropy*. This is based on the fact that, in physics, at least entropy has a quite specific meaning and is defined for systems in thermal equilibrium. When we deal with physical systems far from thermal equilibrium or biological systems, we have to interpret the results anew and we must use new constraints in order to maximize the information or (statistical) entropy.

In particular we saw that, in a certain sense, a new type of information arises, a type of information that refers to the collective variables or order parameters. This suggests that we call the part of information that refers to the order parameters, and which mirrors the collective properties of the system, "synergetic information". At the same time the order parameters adopt a new meaning, namely that of "informators".

It is important to discuss what kind of information, in the usual sense of the word, can be revealed by our general approach based on the use of moments of variables. As we have seen, in this way we can determine the order parameters and their stationary distribution functions. Whether this kind of information is sufficient or not, depends on what we are aiming to achieve. Let us consider the example of a laser. The stationary distribution function is quite sufficient to characterize the steady-state intensity and intensity fluctuations of laser light. But this distribution



function does not tell us anything about a very important property of laser light, namely its temporal coherence. To learn about this one has to deal in much greater detail with the dynamics. In the framework of our approach this kind of dynamics can be guessed from the time-dependent moments and from the transitional probabilities that can be derived from them by means of the maximum information entropy principle. Generally speaking, once averages over specific macroscopic data are known, the whole procedure allows us to guess microscopic events and processes.

It is tempting to conclude that in systems far from thermal equilibrium or even in nonphysical systems, (Shannon) information plays the same role as entropy in systems in thermal equilibrium or close to it, namely as the cause of processes. I am reluctant to follow this interpretation for the following reason: First of all we found that the results of our analysis depend on properly chosen constraints. Here we were able to define adequate constraints for a class of phenomena which can be characterized as nonequilibrium phase transitions. And we may make guesses on the stationary distribution function. In that respect we may say that we can define a kind of potential which drives the system to specific stationary states. But this stationary potential does not give us a unique prescription for deriving for instance the Fokker-Planck equation which determines the dynamics, an aspect stressed by R. Landauer time and again. Thus it seems that the maximization of information (or entropy) is not a fundamental law which drives systems in a unique way. On the other hand there have been very remarkable studies by Graham and Tél and others on the construction of potentials for stochastic processes in situations where bifurcations occur. We have not attempted here to compare our approach with theirs, and this remains an interesting problem for further research. ☆

These ideas are of basic importance for questions concerning biological evolution or development, or in other words, phylogenesis and ontogenesis, of animals. The question, of course, is whether evolution and development are governed by extremal principles, especially extremal principles connected with a single function, such as entropy or information. Our results might be a hint that such a function exists, but the price to be paid is that nothing or only little can be said about the dynamics. But whether such a function really exists, also remains an open question. On the other hand the fact that the choice of moments provided us with the adequate constraints for information in nonequilibrium phase transitions demonstrates the power of our approach. The behavior of the moments reflects the tendency of systems to become coherent. So what we have been studying here is the emergence of coherence and macroscopic order in complex systems via self-organization. But whether biology may be viewed totally from this point of view is another unanswered question. As we said before, complex systems seem to be inexhaustible.

Thus although a deep and fundamental question remains open, the method we have outlined in this book is certainly a powerful tool to study the coherence properties of complex systems close to nonequilibrium phase transitions from a unifying point of view.