vector with components $q_1, \ldots, q_m$ the Îto equation reads

$$dq_l(t) = K_l(\boldsymbol{q}(t))\,dt + \sum_m g_{lm}(\boldsymbol{q}(t))\,dw_m(t) \tag{2.12}$$

where dw_m, which describes the stochastic process, has the properties

$$\langle dw_m \rangle = 0 \tag{2.13}$$

$$\langle dw_m(t)\,dw_l(t') \rangle = \delta_{lm}\delta(t - t')\,dt \ . \tag{2.14}$$

To exhibit the special features of the Îto procedure let us consider an arbitrary differentiable function

$$u = u(\boldsymbol{q}) \tag{2.15}$$

and its differential. Because of the property (2.14) we have to go up to the second derivative according to

$$du_j = \sum_k \frac{\partial u_j}{\partial q_k} dq_k + \frac{1}{2} \sum_{kl} \frac{\partial^2 u_j}{\partial q_k \partial q_l} dq_k\,dq_l \ . \tag{2.16}$$

Inserting (2.12) into (2.16) and keeping the terms including dt but neglecting all higher order terms we arrive at

$$du_j = \sum_k \frac{\partial u_j}{\partial q_k} [K_k(q)\,dt + \sum_m g_{km}\,dw_m(t)] + \frac{1}{2} \sum_{kl} \frac{\partial^2 u_j}{\partial q_k \partial q_l} \left[\sum_{mn} g_{km} g_{ln}\,dw_m\,dw_n \right] . \tag{2.17}$$

Let us now briefly remind the reader of the *Stratonovich approach*. The stochastic equation for a single variable reads

$$dq = K(q)\,dt + g(q)\,dw(t) \tag{2.18}$$

and for the components of a state vector

$$dq_l = K_l(\boldsymbol{q})\,dt + \sum_m g_{lm}(\boldsymbol{q})\,dw_m(t) \ . \tag{2.19}$$

In contrast to the Îto calculus, the last terms in (2.18) or (2.19) are now interpreted differently, namely they have to be evaluated according to the midpoint rule in which $g(q(t_i))\,dw(t_i)$ is replaced by

$$g\left(q\left(\frac{t_i + t_{i-1}}{2}\right)\right) dw(t_i) \ , \tag{2.20}$$

i.e. q and dw are no more statistically independent. As we shall see later, we can recover the Îto equation by means of a macroscopic approach.

2.3 Fokker-Planck Equation

For many applications, especially when the problems are non-linear, i.e. when K is a non-linear function of q, it is advantageous to proceed to the Fokker-Planck equation which is formulated for the distribution function $f(q,t)$. It describes the probability of finding the variable q in the interval $q \to q + dq$ at time t.

The Fokker-Planck equation belonging to the Langevin equation (2.1) reads

$$\frac{\partial f}{\partial t} = -\frac{\partial}{\partial q}(K(q)f) + \frac{Q}{2}\frac{\partial^2}{\partial q^2}f \tag{2.21}$$

where the first term on the right-hand side is denoted as the drift term, and the second term is called the diffusion term.

The stationary solution obeying

$$\frac{\partial f}{\partial t} = 0 \tag{2.22}$$

can easily be found, provided the boundary condition

$$f(q) \to 0 \qquad \text{for} \qquad q \to \mp\infty \tag{2.23}$$

is fulfilled. Then the stationary solution reads

$$f = N \exp\left[-\int_{q_0}^{q} \frac{2K(q')}{Q} dq' \right] \tag{2.24}$$

where N is the normalization factor so that the integral over f is equal to unity. In the case of a multidimensional state vector, we have to find the Fokker-Planck equation for

$$f(\boldsymbol{q}, t) \ . \tag{2.25}$$

This equation reads

$$\frac{\partial f}{\partial t} = -\sum_j \frac{\partial}{\partial q_j}(K_j f) + \frac{1}{2}\sum_{jk} Q_{jk}\frac{\partial^2}{\partial q_j \partial q_k} f \ . \tag{2.26}$$

Explicit solutions of (2.26) are available only in special cases, e.g. if K is linear in the variables q and Q_{jk} is independent of $\boldsymbol{q}$. In such a case the time dependent and stationary solutions can be constructed explicitly (cf. my book *Advanced Synergetics*). The stationary solution of (2.26) can also be constructed explicitly, provided the so-called rule of detailed balance is fulfilled (see below, Sect. 2.4).

When we start from the *Îto differential equation*, the corresponding Fokker-Planck equation has the following form

$$\frac{\partial f}{\partial t} = -\sum_k \frac{\partial}{\partial q_k}[K_k(\boldsymbol{q})f] + \frac{1}{2}\sum_{kl} \frac{\partial^2}{\partial q_k \partial q_l}\left[\sum_m g_{km} g_{lm} f \right] . \tag{2.27}$$

In the case of the *Stratonovich calculus*, however, the Fokker-Planck equation can be shown to read

$$\frac{\partial f}{\partial t} = -\sum_l \frac{\partial}{\partial q_l}\left\{\left[K_l(q) + \frac{1}{2}\sum_{kj}\frac{\partial g_{lj}}{\partial q_k} g_{kj}\right] f\right\} + \frac{1}{2}\sum_{lm}\frac{\partial^2}{\partial q_l \partial q_m}\left(\sum_i g_{li} g_{mi} f\right) . \tag{2.28}$$

2.4 Exact Stationary Solution of the Fokker-Planck Equation for Systems in Detailed Balance

In this section we show that under the condition of detailed balance the stationary solution of the Fokker-Planck equation may be found explicitly by quadratures.

While the principle of detailed balance is expected to hold for practically all systems in thermal equilibrium, this need not be so in systems far from thermal equilibrium. Thus each individual case requires a detailed discussion (e.g., by symmetry considerations) as to whether this principle is applicable. Also, an inspection of the structure of the Fokker-Planck equation will enable us to decide whether detailed balance is present.

2.4.1 Detailed Balance

We denote the set of variables $q_1, \ldots, q_N$ by $\boldsymbol{q}$ and the set of the variables under time reversal by

$$\tilde{\boldsymbol{q}} = \{\varepsilon_1 q_1, \ldots, \varepsilon_N q_N\} , \tag{2.29}$$

where $\varepsilon_i = -1\,(+1)$ depending on whether the coordinate q_i changes sign (does not change sign) under time reversal. Furthermore, λ stands for a set of externally determined parameters. The time reversed quantity is denoted by

$$\tilde{\boldsymbol{\lambda}} = \{\nu_1 \lambda_1, \ldots, \nu_M \lambda_M\} , \tag{2.30}$$

where $\nu_i = -1\,(+1)$ depends on the inversion symmetry of the external parameters under time reversal. We denote the joint probability of finding the system at t_1 with coordinates $\boldsymbol{q}$ and at t_2 with coordinates $\boldsymbol{q}'$ by

$$P(\boldsymbol{q}', \boldsymbol{q}; t_2, t_1) . \tag{2.31}$$

In the following, we consider a stationary system so that the joint probability depends only on the time difference $t_2 - t_1 = \tau$. Thus (2.31) may be written as

$$P(\boldsymbol{q}', \boldsymbol{q}; t_2, t_1) = W(\boldsymbol{q}', \boldsymbol{q}; \tau) . \tag{2.32}$$

We now formulate the principle of detailed balance. The following two definitions are available.

1) The principle of detailed balance (first version)

$$W(\boldsymbol{q}', \boldsymbol{q}; \tau, \lambda) = W(\tilde{\boldsymbol{q}}, \tilde{\boldsymbol{q}}'; \tau, \tilde{\lambda}) \ . \tag{2.33}$$

The joint probability may be expressed by the stationary distribution $f(\boldsymbol{q})$ multiplied by the conditional probability P, where stationarity is exhibited by writing

$$P = P(\boldsymbol{q}'|\boldsymbol{q}; \tau, \lambda) \ . \tag{2.34}$$

Therefore, we may reformulate (2.33) as follows:

$$P(\boldsymbol{q}'|\boldsymbol{q}; \tau, \lambda) f(\boldsymbol{q}, \lambda) = P(\tilde{\boldsymbol{q}}|\tilde{\boldsymbol{q}}'; \tau, \tilde{\lambda}) f(\tilde{\boldsymbol{q}}', \tilde{\lambda}) \ . \tag{2.35}$$

Here and in the following we assume that the Fokker-Planck equation possesses a unique stationary solution. One may then show directly that

$$f(\boldsymbol{q}, \lambda) = f(\tilde{\boldsymbol{q}}, \tilde{\lambda}) \tag{2.36}$$

holds. We define the transition probability per second by

$$w(\boldsymbol{q}', \boldsymbol{q}; \lambda) = [(d/d\tau) P(\boldsymbol{q}'|\boldsymbol{q}; \tau, \lambda)]_{\tau=0} \ . \tag{2.37}$$

Taking the derivative with respect to τ on both sides of (2.35) and putting $\tau = 0$ (but $\boldsymbol{q} \neq \boldsymbol{q}'$), we obtain

2) the principle of detailed balance (second version)

$$w(\boldsymbol{q}', \boldsymbol{q}; \lambda) f(\boldsymbol{q}, \lambda) = w(\tilde{\boldsymbol{q}}, \tilde{\boldsymbol{q}}'; \tilde{\lambda}) f(\tilde{\boldsymbol{q}}', \tilde{\lambda}) \ . \tag{2.38}$$

This obviously has a very simple meaning. The left-hand side describes the total transition rate out of the state $\boldsymbol{q}$ into a new state $\boldsymbol{q}'$. The principle of detailed balance then requires that this transition rate is equal to the rate in the reverse direction for $\boldsymbol{q}'$ and $\boldsymbol{q}$ with reverse motion, e.g., with reverse momenta. It can be shown that the first and second version are equivalent.

2.4.2 The Required Structure of the Fokker-Planck Equation and Its Stationary Solution

Using the conditional probability P (which is nothing but the Green's function) we write the Fokker-Planck equation in the form

$$\frac{d}{d\tau} P(\boldsymbol{q}'|\boldsymbol{q}; \tau, \lambda) = L(\boldsymbol{q}', \lambda) P(\boldsymbol{q}'|\boldsymbol{q}; \tau, \lambda) \tag{2.39}$$

where we assume that the operator L has the form

$$L(\boldsymbol{q}) = -\sum_i \frac{\partial}{\partial q_i} K_i(\boldsymbol{q}, \lambda) + \frac{1}{2} \sum_{ik} \frac{\partial^2}{\partial q_i \partial q_k} Q_{ik}(\boldsymbol{q}, \lambda) \ . \tag{2.40}$$

We may always assume that the diffusion coefficients are symmetric

$$Q_{ik} = Q_{ki} \ . \tag{2.41}$$

We define the following new coefficients:

a) the irreversible drift coefficients

$$D_i(q, \lambda) = \tfrac{1}{2}[K_i(q, \lambda) + \varepsilon_i K_i(\tilde{q}, \tilde{\lambda})] \equiv D_i^{ir} \ ; \tag{2.42}$$

b) the reversible drift coefficients

$$J_i(q, \lambda) = \tfrac{1}{2}[K_i(q, \lambda) - \varepsilon_i K_i(\tilde{q}, \tilde{\lambda})] \equiv D_i^{r} \ . \tag{2.43}$$

J_i transforms as q_i under time reversal.

We write the stationary solution of the Fokker-Planck equation in the form

$$f(q, \lambda) = \mathcal{N} \mathrm{e}^{-\Phi(q, \lambda)} \ , \tag{2.44}$$

where $\mathcal{N}$ is the normalization constant and Φ may be interpreted as a generalized thermodynamic potential. The necessary and sufficient conditions for the principle of detailed balance to hold read

$$Q_{ik}(q, \lambda) = \varepsilon_i \varepsilon_k Q_{ik}(\tilde{q}, \tilde{\lambda}) \ , \tag{2.45}$$

$$D_i - \frac{1}{2} \sum_k \frac{\partial Q_{ik}}{\partial q_k} = -\frac{1}{2} \sum_k K_{ik} \frac{\partial \Phi}{\partial q_k} \ , \tag{2.46}$$

$$\sum_i \left(\frac{\partial J_i}{\partial q_i} - J_i \frac{\partial \Phi}{\partial q_i} \right) = 0 \ . \tag{2.47}$$

If the diffusion matrix Q_{ik} possesses an inverse, (2.46) may be solved for the gradient of Φ

$$\frac{\partial \Phi}{\partial q_i} = \sum_k (Q^{-1})_{ik} \left(\sum_l \frac{\partial Q_{kl}}{\partial q_l} - 2D_k \right) \equiv A_i \ . \tag{2.48}$$

This shows that (2.48) implies the integrability condition

$$\frac{\partial}{\partial q_j} A_i = \frac{\partial}{\partial q_i} A_j \ , \tag{2.49}$$

which is a condition on the drift and diffusion coefficients as defined by the right-hand side of (2.48). Substituting A_i and A_j from (2.48), the condition (2.47) acquires the form

$$\sum_i \left[\frac{\partial J_i}{\partial q_i} - J_i \sum_k (Q^{-1})_{ik} \left(\sum_l \frac{\partial Q_{kl}}{\partial q_l} - 2D_k \right) \right] = 0 \ . \tag{2.50}$$

Thus the conditions for detailed balance to hold are given finally by (2.45,49,50). Equation (2.46) or equivalently (2.48) then allows us to determine Φ by pure

quadratures, i.e., by a line integral. Thus the stationary solution of the Fokker-Planck equation may be determined explicitly.

2.5 Path Integrals

The time-dependent solutions of the Fokker-Planck equation can be represented in the form of path integrals. For the sake of simplicity we shall be concerned in the following with the case where the diffusion constant Q is independent of $\boldsymbol{q}$. Let us first treat the one dimensional case, i.e. that $\boldsymbol{q}$ is a single variable. Let us split the time interval t into equidistant steps

$$t_0, t_1 = t_0 + \tau, \ldots, t_N = t_0 + \tau N \ . \tag{2.51}$$

The distribution function f at time t can be constructed as a multiple integral over all the intermediate positions $q_0, q_1, q_2, \ldots$ (cf. Fig.2.2). The explicit form of the path integral then reads

$$f(q,t) = \lim_{\substack{N\to\infty \\ N\tau=t}} \int \cdots \int Dq\, \mathrm{e}^{-G/2} f(q', t_0) \tag{2.52}$$

where we have used the abbreviations

$$Dq = (2Q\tau\pi)^{-N/2}\, dq_0, \ldots, dq_{N-1} \qquad \text{and} \tag{2.53}$$

$$G = \sum_{\nu} \frac{\tau\left[\dfrac{q_\nu - q_{\nu-1}}{\tau} - K(q_{\nu-1})\right]^2}{Q} \ . \tag{2.54}$$

The explicit derivation of formula (2.52) is presented in *Synergetics. An Introduction.* Let us now consider the generalization of (2.52) to the case of a multi-dimensional vector $\boldsymbol{q}$ which has n components. We still assume that Q_{mn} is independent of the state variable $\boldsymbol{q}$. We then obtain the following results

$$f(\boldsymbol{q},t) = \lim_{\substack{N\to\infty \\ N\tau=t}} \int \cdots \int Dq\, \mathrm{e}^{-G/2} f(\boldsymbol{q}', t_0) \tag{2.55}$$

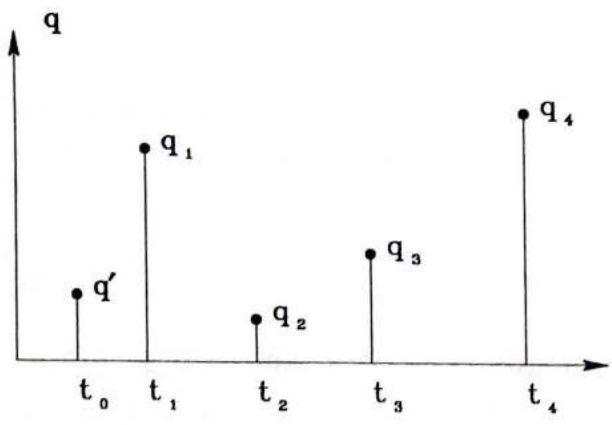

Fig. 2.2. The positions of a particle or of the state of a system in the course of time

where we have used the following abbreviations

$$Dq = \prod_{\mu=0}^{N-1} \{(2\pi\tau)^{-n/2}(\det Q)^{-1/2}\}(dq_1, \ldots, dq_n)_\mu \tag{2.56}$$

$$q_N = q \ ; \qquad q_0 = q' \tag{2.57}$$

$$G = \tau \sum (\dot{q}_\nu^T - K_{\nu-1}^T) Q^{-1} (\dot{q}_\nu - K_{\nu-1}) \tag{2.58}$$

$$\dot{q}_\nu = \tau^{-1}(q_\nu - q_{\nu-1}) \ , \qquad K_{\nu-1} = K(q_{\nu-1}) \ , \tag{2.59}$$

where T denotes the transposed vector.

Let us finally remind the reader of the master equation. Let us consider a discrete state space vector m. Then we are interested in the probability distribution $P(m, t)$. Provided we are dealing with a Markovian process, P obeys the master equation.

$$\frac{dP(m,t)}{dt} = \sum_n w(m,n)P(n) - \sum_n w(n,m)P(m) \ . \tag{2.60}$$

Again, it is difficult to find explicit solutions of (2.60). In the special case of detailed balance, the stationary probability distribution can be constructed explicitly. When detailed balance is present, P fulfills the following relation

$$w(n,m)P(m) = w(m,n)P(n) \ . \tag{2.61}$$

Then the steady state solution of (2.60) can be written down explicitly in the form

$$P(m) = N\,\mathrm{e}^{\Phi(m)} \tag{2.62}$$

where $\Phi(m)$ is defined by

$$\Phi(m) = \Phi(n_0) + \sum_{j=0}^{N-1} \ln\left\{\frac{w(n_{j+1}, n_j)}{w(n_j, n_{j+1})}\right\} \qquad \text{and} \tag{2.63}$$

$$m \equiv n_N \ . \tag{2.64}$$

2.6 Reduction of Complexity, Order Parameters and the Slaving Principle

In this section we treat systems which are composed of many parts. We wish to study qualitative changes in the behavior of the system. To this end we make several transformations of the variables and their equations. Then in Sects. 2.7,8 we shall present some important applications. We start from a state vector q which describes the total system at the microscopic or mesoscopic level.

$$q = q(x, t) \ . \tag{2.65}$$

For what follows we shall assume that the state vector is a function of the space coordinate x so that quite generally we shall assume that q obeys an evolution equation of the type

$$\dot{q} = N(q, \alpha) + F(t) \ . \tag{2.66}$$

In this equation N is a nonlinear function of q that may also contain differential operators, for instance the Laplace operator differentiating q which respect to the spatial coordinates. α is a control parameter, e.g. the power input into a laser, or the amount of heating of a fluid, or certain signals impinging on a biological system, and $F(t)$ is a fluctuating force. We now proceed in several steps.

2.6.1 Linear Stability Analysis

In the following we shall assume that for a fixed value of the control parameter, α_0, the solution of the deterministic equation is known, i.e. that q_0 solves the equation

$$\dot{q} = N(q, \alpha_0) \ . \tag{2.67}$$

We then study the behavior of the solution when the control parameter α is changed. To this end we make the hypothesis

$$\alpha: q = q_0 + w \ . \tag{2.68}$$

We assume that q_0 changes smoothly with α

$$q_0 = q_0(\alpha) \ . \tag{2.69}$$

We wish to study the stability of that solution q_0. We thus insert the hypothesis (2.68) into the equation (2.66) where F however, is dropped. We then obtain

$$\dot{q}_0 + \dot{w} = N(q_0 + w, \alpha) \ . \tag{2.70}$$

Under the assumption that w is a small quantity, we may expand the right-hand side of (2.70) into a power series in w and keep only the two leading terms

$$\dot{q}_0 + \dot{w} = N(q_0, \alpha) + L(q_0)w + \cdots \ . \tag{2.71}$$

On account of (2.67), the first term on the l.h.s. of (2.70) cancels with the first term on the r.h.s. of (2.71) so that we are left with the equation

$$\dot{w} = L(q_0)w \ . \tag{2.72}$$

Note that L, which depends on q_0, may still contain differential operators acting on the space coordinates in w. Nevertheless the general solution of (2.72) can be written in the form

$$w(t) = e^{\lambda t} v \ . \tag{2.73}$$

where v is a time-independent vector.

We note that more general cases have also been treated, namely those where q_0 is a periodic or quasiperiodic function. For the detailed results I must refer the reader to my book *Advanced Synergetics*. For what follows it is important to distinguish between the so-called unstable and stable modes. The unstable modes are those for which

$$\lambda > 0 \tag{2.74}$$

holds. They shall be denoted by

$$\lambda_u, v_u \ . \tag{2.75}$$

The stable modes are characterized by

$$\lambda < 0 \tag{2.76}$$

and shall be denoted by

$$\lambda_s, v_s \ . \tag{2.77}$$

Note that the terms "unstable" and "stable" refer only to the linear analysis. In fact, it will turn out that in general the so-called unstable modes will become stabilized by means of their interaction with the stable modes. Note further, that our approach is a fully nonlinear one and the linear stability analysis serves only to find an adequate frame of reference in which to represent the desired solution q of (2.66).

2.6.2 Transformation of Evolution Equations

In order to solve (2.66) in the nonlinear and stochastic case, we make the hypothesis

$$q = q_0 + \sum_u \xi_u(t) v_u + \sum_s \xi_s(t) v_s \ . \tag{2.78}$$

In the case where L contains differential operators acting on space variables, v is a space dependent function

$$v = v(x) \ . \tag{2.79}$$

When we insert (2.78) into (2.66) and project both sides onto the expansion functions v_u and v_s, we obtain equations for the mode amplitudes ξ_u and ξ_s

$$\dot{\xi}_u = \lambda_u \xi_u + N_u(\xi_u, \xi_s) + F_u(t) \tag{2.80}$$

$$\dot{\xi}_s = \lambda_s \xi_s + N_s(\xi_u, \xi_s) + F_s(t) \ . \tag{2.81}$$

The indices u and s serve two purposes. Firstly, they indicate whether we are dealing with the amplitudes of the unstable or of the stable modes. Secondly, they serve to number the individual components of ξ_u and ξ_s. For instance we may let u and s take the values $u = 1, \ldots, M$ and $s = M + 1, \ldots$. The context will show which meaning the indices u or s have to be given. The amplitudes ξ_u will be called order parameters.

2.6.3 The Slaving Principle

The transformation (2.78) of the equation (2.66) does not reduce the complexity, i.e. the equations (2.80) and (2.81) are fully equivalent to the equations (2.66). The slaving principle of synergetics allows us, however, to eliminate from (2.80) and (2.81) the slaved mode amplitudes by means of an explicit formula

$$\xi_s(t) = f_s[\xi_u(t), t] \ . \tag{2.82}$$

The explicit construction of f_s is described in my book *Advanced Synergetics* and in special cases also in my book *Synergetics. An Introduction.* Here, we just illustrate the contents of (2.82) by means of a simple explicit example where we present the slaving principle in its leading term. To this end let us consider the equations for the amplitudes ξ_u and ξ_s, namely (2.80,81), in the following form

$$\dot{\xi}_u = \lambda_u \xi_u + h_u(\xi_u, \xi_s) + F_u(t) \tag{2.83a}$$

$$\dot{\xi}_s = \lambda_s \xi_s + g_s(\xi_u) + q_s k_s(\xi_u) + F_s(t) \ . \tag{2.83b}$$

Here it is assumed that h_u is a nonlinear function which starts with powers of at least second order in ξ_u. Similarly, g_s is a function starting at the same power. It may then be shown that ξ_s starts with powers of at least second order in ξ_u. In its simplest form, the slaving principle amounts to putting $\dot{\xi}_s$ in (2.83b) equal to zero.

Keeping the leading orders we readily obtain the result

$$\xi_s \approx -\frac{1}{\lambda_s} g_s(\xi_u) - \frac{1}{\lambda_s} F_s(t) \ . \tag{2.84}$$

This result can be proven rigorously to lowest order in F_s and ξ_u. We now wish to study what the slaving principle means for the solution of the Fokker-Planck equation. For this purpose we transform the Fokker-Planck equation from the old state vector q into the new variables ξ_u, ξ_s

$$q \to \xi_u, \xi_s \ . \tag{2.85}$$

The Fokker-Planck equation then acquires the general form

$$\dot{P}(\xi_u, \xi_s; t) = L(\xi_u, \xi_s) P(\xi_u, \xi_s; t) \tag{2.86}$$

where L is a linear operator. Let us now consider the steady state solution of (2.86). It can always be written in the form

$$P(\xi_u, \xi_s) = P(\xi_s | \xi_u) f(\xi_u) \tag{2.87}$$

where the l.h.s. is a joint probability whereas P on the r.h.s. is a conditional probability. f is a distribution function for the order parameters alone. The slaving principle, in its leading approximation, now means that the conditional probability P on the r.h.s. of (2.87) can be written more specifically as

$$P(\xi_s | \xi_u) = \prod_s P_s(\xi_s | \xi_u) \ . \tag{2.88}$$

Our result (2.84) can now be used to give us an explicit example of what (2.88) may look like in this lowest order approximation of the slaving principle. The fluctuating forces are, as usual, Gaussian distributed, i.e. the probability of finding F_s within the interval $F \to F + dF$ is given by

$$P(F \leq F_s \leq F + dF) = N' \exp(-F_s^2/Q')\, dF \ . \tag{2.89}$$

Now we may solve the relation (2.84) for F_s

$$F_s = -\lambda_s \left[\xi_s + \frac{1}{\lambda_s} g_s(\xi_u) \right] . \tag{2.90}$$

This allows us to determine the conditional probability by using (2.89). In this way we obtain

$$P(\xi_s|\xi_u) = N \exp\left\{ -\left[\xi_s + \frac{1}{\lambda_s} g_s(\xi_u) \right]^2 \Big/ Q \right\} d\xi_s \ . \tag{2.91}$$

where we have used the abbreviations

$$Q^{-1} = Q'^{-1} \lambda_s^2 \tag{2.92}$$

$$N = N' \lambda_s \ . \tag{2.93}$$

2.7 Nonequilibrium Phase Transitions

In many cases of practical interest, the number of order parameters may be very small or even one, whereas the number of slaved modes is still very large. Let us consider the case of a single order parameter and let us drop the index u for simplicity

$$\xi_u \to \xi, F_u \to F, \lambda_u \to \lambda \ . \tag{2.94}$$

A typical order parameter equation then reads

$$\dot{\xi} = \lambda \xi - \beta \xi^3 + F(t) \ , \tag{2.95}$$

as is shown in synergetics.

The fluctuating force $F(t)$ obeys the relation

$$\langle F(t) F(t') \rangle = Q \delta(t - t') \ . \tag{2.96}$$

If we assume in addition that F is Gaussian distributed we may establish a Fokker-Planck equation belonging to (2.95) in the form

$$\dot{f}(\xi; t) = -\frac{\partial}{\partial \xi} [(\lambda \xi - \beta \xi^3) f] + \frac{Q}{2} \frac{\partial^2}{\partial \xi^2} f \ . \tag{2.97}$$

Let us consider the steady state for which

$$\dot{f} = 0 \ . \tag{2.98}$$

Then (2.97) can easily be integrated,

$$f = N \exp[Q^{-1}(\lambda \xi^2 - \tfrac{1}{2}\beta \xi^4)] \ . \tag{2.99}$$

Equation (2.99) provides us with an explicit and typical example for the distribution function f occurring in (2.87). Distribution functions for order parameters can be found also for several order parameters explicitly, provided the principle of detailed balance holds. Let us assume that λ_u is real. A special case of the principle of detailed balance is the following: The Langevin equation has the form

$$\dot{\xi}_u = \lambda_u \xi_u + \frac{\partial V}{\partial \xi_u} + F_u(t) \tag{2.100}$$

where V is a nonlinear function of ξ_u. We assume that

$$\langle F_u(t) F_{u'}(t') \rangle = \delta_{uu'} Q \delta(t - t') \tag{2.101}$$

holds. Then the solution of the Fokker-Planck equation belonging to the Langevin equation (2.100) can be written in the general form

$$f(\xi_u) = N \exp\left[-\frac{2V(\xi_u)}{Q} \right] . \tag{2.102}$$

Even if the principle of detailed balance is not valid, but the distribution function is singly connected, it may always be written in the form

$$f(\xi_u) = N \exp[-\Phi(\xi_u)] \tag{2.103}$$

where Φ plays the role of a generalized thermodynamic potential. Now let us assume that a soft transition occurs when we change the control parameter so that the system becomes unstable and enters a new region. In such a case we may assume that the order parameters are still small, for instance that

$$\xi_u \propto \lambda_u^{\kappa} \ , \quad \kappa > 0 \tag{2.104}$$

In such a case we may expand Φ as a power series with respect to ξ_u

$$\Phi(\xi_u) = \sum_u \lambda_u \xi_u^2 + \sum_{uu'} c_{uu'} \xi_u \xi_{u'} + \sum_{uu'u''} c_{uu'u''} \xi_u \xi_{u'} \xi_{u''} + \sum_{uu'u''u'''} c_{uu'u''u'''} \xi_u \xi_{u'} \xi_{u''} \xi_{u'''} \ . \tag{2.105}$$

In most cases of non-equilibrium phase transitions, it is sufficient to retain the first 4 powers, though in exceptional cases higher powers are also needed. In general,

systems possess internal symmetries that give rise to specific relations between the coefficients c of each order.

2.8 Pattern Formation

We now wish to show how our above formalism can describe the formation of patterns. If we are dealing with continuously extended media which are described by a space coordinate x, then in general, the operator L in the linearization (2.72) contains derivatives with respect to the spatial coordinate. In such a case w and thus v which occurs in (2.73) become functions of the space coordinate x, (2.79).

Quite generally the solution q of the nonlinear equation (2.66) can be written in the form

$$q = q_0 + \sum_u \xi_u(t) v_u(x) + \sum_s \xi_s(t) v_s(x) \ . \tag{2.106}$$

which is just the same as our previous formula (2.78).

As it turns out, ξ_u is, in general, an order of magnitude larger than ξ_s, i.e. the evolving pattern is mainly determined by the first sum over u in (2.106) which we therefore call the mode skeleton. If only a single order parameter is present and v_u has the form

$$v_u = L^{-1/2} \sin kx \ , \tag{2.107}$$

then (2.106) is essentially given by

$$q = q_0 + \xi_u(t) L^{-1/2} \sin kx \ . \tag{2.108}$$

In many cases ξ_u obeys an equation of the form (2.95) which describes the growth of ξ_u out of an initial fluctuation to its final size (Fig.2.3,4). Quite evidently, with more order parameters and/or more complicated functions v_u, far more complicated patterns than that described by (2.107), and (2.108) can be obtained. Thus, this theory is capable of deriving the emergent spatial structures of complex systems.

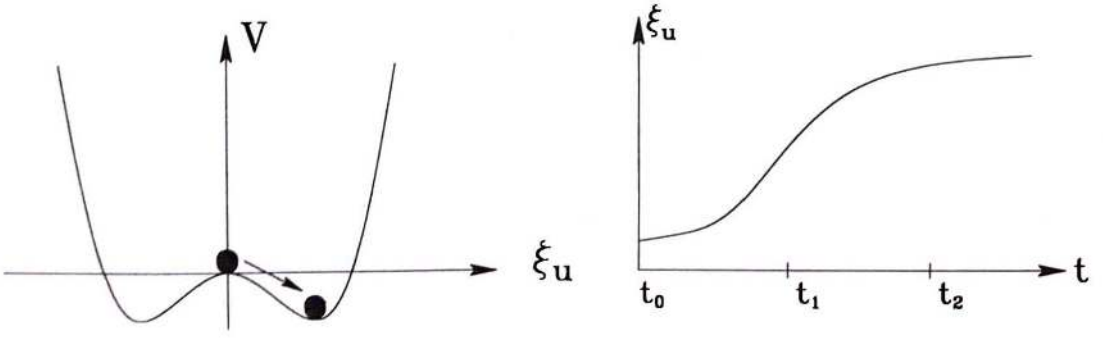

Fig. 2.3. Illustration of the behavior of the order parameter ξ_u as a function of time. Left-hand side: The potential function V in which a fictitious particle with coordinate ξ_u may move. Right-hand side: The temporal evolution of ξ_u

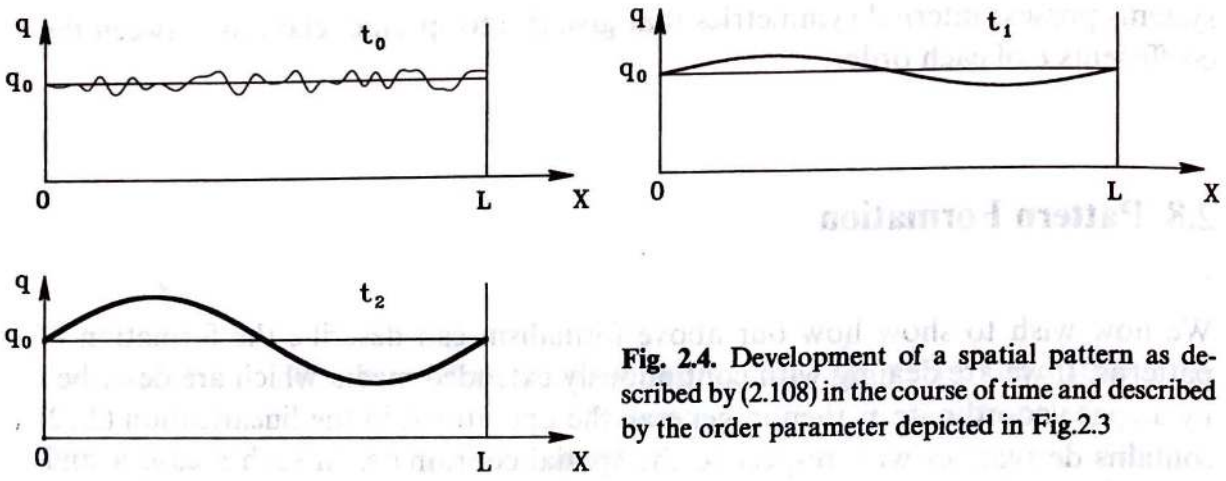

Fig. 2.4. Development of a spatial pattern as described by (2.108) in the course of time and described by the order parameter depicted in Fig.2.3

Our example here is just a brief reminder of what has been presented in my books *Synergetics* and *Advanced Synergetics*. I just wish to remind the reader that considerably more general cases have been treated there, such as evolving time-dependent patterns, namely limit cycles, quasi periodic motion, or chaos. But for what follows this brief reminder will, in most cases, be sufficient.

3. ... and Back Again: The Maximum Information Principle (MIP)

3.1 Some Basic Ideas

In this chapter we address the following question: Let some *macroscopic quantities* of a system be given. We then wish to devise a procedure by which we can derive the probability distribution of *macroscopic* or even *microscopic* variables. In other words, we start from the macroscopic world and wish to draw conclusions about the microscopic world. Depending on the kind of systems we are treating, the adequate macroscopic quantities may be quite different. In closed physical systems, to which thermodynamics applies, these quantities are energy, particle numbers etc., and we shall illustrate the general procedure by this example in Chap. 4. In open systems, e.g. in physics or biology, the adequate macroscopic quantities will turn out to be, for instance, intensities and intensity fluctuations. Indeed, it will be the main purpose of the following chapters, to deal with *open systems.*

Since the starting point of our approach is the concept of information, we shall derive this concept in this Sect. 3.1.

By some sort of new interpretation of probability theory we get an insight into a seemingly quite different discipline, namely information theory. Consider the sequence of tossing a coin with outcomes 0 and 1. Now interpret 0 and 1 as a dash and dot of a Morse alphabet. We all know that by means of a Morse alphabet we can transmit messages so that we may ascribe a certain meaning to a certain sequence of symbols. Or, in other words, a certain sequence of symbols carries information. In information theory we try to find a measure for the amount of information.

Let us consider a simple example and consider R_0 different possible events ("realizations") which are equally probable a priori. Thus when tossing a coin we have the events 1 and 0 and $R_0 = 2$. In the case of a die we have 6 different outcomes, therefore $R_0 = 6$. Thus the outcome of tossing a coin or throwing a die is interpreted as the receipt of a message and only one out of the possible R_0 outcomes is actually realized. Apparently the greater R_0, the greater is the uncertainty before the message is received and the larger will be the amount of information after the message is received. Thus we may interpret the whole procedure in the following manner: In the initial situation we have no information I_0, i.e., $I_0 = 0$ with R_0 equally probable outcomes.

In the final situation we have an information $I_1 \neq 0$ with $R_1 = 1$, i.e., a single outcome. We now want to introduce a measure for the amount of information, I, which apparently must be connected with R_0. To get an idea how the connection between R_0 and I must appear we require that I is additive for independent events.

Thus when we have two such sets with R_{01} or R_{02} outcomes so that the total number of outcomes is

$$R_0 = R_{01} R_{02} \tag{3.1}$$

we require

$$I(R_{01} R_{02}) = I(R_{01}) + I(R_{02}) \ . \tag{3.2}$$

This relation can be fulfilled by choosing

$$I = K \ln R_0 \tag{3.3}$$

where K is a constant. It can even be shown that (3.3) is the only solution to (3.2). The constant K is still arbitrary and can be fixed by some definition. Usually the following definition is used. We consider a so-called "binary" system which has only two symbols (or letters). These may be the head and the tail of a coin, or answers yes and no, or numbers 0 and 1 in a binomial system. When we form all possible "words" (or sequences) of length n, we find $R = 2^n$ realizations. We now want to identify I with n in such a binary system. We therefore require

$$I = K \ln R \equiv Kn \ln 2 = n \tag{3.4}$$

which is fulfilled by

$$K = \frac{1}{\ln 2} = \log_2 \mathrm{e} \tag{3.5}$$

With this choice of K, another form of (3.4) reads

$$I = \log_2 R \ . \tag{3.4a}$$

Since a single position in a sequence of symbols (signs) in a binary system is called "bit", the information I is now directly given in bits. Thus if $R = 8 = 2^3$ we find $I = 3$ bits and generally for $R = 2^n$, $I = n$ bits. The definition of information for (3.3) can be easily generalized to the case where we initially have R_0 equally probable cases and finally R_1 equally probable cases. In this case the information is

$$I = K \ln R_0 - K \ln R_1 \tag{3.6}$$

which reduces to the earlier definition (3.3), if $R_1 = 1$. A simple example for this is given by a die. Let us define a game in which the even numbers mean gain and the odd numbers mean loss. Then $R_0 = 6$ and $R_1 = 3$. In this case the information content is the same as that of a coin with originally just two possibilities.

We now derive a more convenient expression for the information: We first consider the following example of a simplified Morse alphabet with dash and dot (in the real Morse alphabet, the intermission is a third symbol). We consider a word of length G which contains N_1 dashes and N_2 dots, with

$$N_1 + N_2 = N \ . \tag{3.7}$$

We ask for the information which is obtained by the receipt of such a word. In the spirit of information theory we must calculate the total number of words which can be constructed out of these two symbols for fixed N_1, N_2. The analysis is quite simple. According to the ways in which we can distribute the dashes and dots over the N positions, there are

$$R = \frac{N!}{N_1!N_2!} \tag{3.8}$$

possibilities. Or, in other words, R is the number of messages which can be transmitted by N_1 dashes and N_2 dots. We now want to derive the information per symbol, i.e. $i = I/N$. Inserting (3.8) into (3.3) we obtain

$$I = K \ln R = K[\ln N! - \ln N_1! - \ln N_2!] \ . \tag{3.9}$$

Using Stirling's formula in the approximation

$$\ln N! \approx N(\ln N - 1) \ , \tag{3.10}$$

which is good for $N > 100$, we readily find

$$I \approx K[N(\ln N - 1) - N_1(\ln N_1 - 1) - N_2(\ln N_2 - 1)] \ , \tag{3.11}$$

and from (3.7) we then have

$$i \equiv \frac{I}{N} \approx -K\left[\frac{N_1}{N}\ln\frac{N_1}{N} + \frac{N_2}{N}\ln\frac{N_2}{N}\right] . \tag{3.12}$$

We now introduce a quantity which may be interpreted as the probability of finding the sign "dash" or "dot". The probability is identical to the relative frequency with which dash or dot are found

$$p_j = \frac{N_j}{N} \ , \qquad j = 1, 2 \ . \tag{3.13}$$

With this, our final formula takes the form

$$i = \frac{I}{N} = -K(p_1 \ln p_1 + p_2 \ln p_2) \ . \tag{3.14}$$

This expression can be easily generalized to the case where we do not simply have two symbols but several, such as letters in the alphabet. Then we obtain, in an exactly analogous manner, an expression for the information per symbol which is given by

$$i = -K \sum_j p_j \ln p_j \ . \tag{3.15}$$

p_j is the relative frequency of the occurrence of the symbols. From this interpretation it is evident that i may be used in the context of transmission of information, etc.

Before continuing we should say a word about information used in the sense here. It should be noted that "useful" or "useless" or "meaningful" or "meaningless" are not contained in the theory; e.g., in the Morse alphabet defined above quite a number of words might be meaningless. Information in the sense used here rather refers to the scarcity of an event. Though this may seem to restrict the theory considerably, the theory will in fact turn out to be extremely useful.

The expression for the information can be viewed in two completely different ways. On the one hand we may assume that the p_i's are given by their numerical values, and then we may write down a number for I by use of formula (3.3). Of still greater importance, however, is a second interpretation; namely, to consider I as a function of the p_i's such that if we change the values of the p_i's, the value of I changes correspondingly. To make this interpretation clear we anticipate an application which we will treat later in much greater detail. Consider a gas of atoms moving freely in a box. It is then of interest to know about the spatial distribution of the gas atoms. We divide the container into M cells of equal size and denote the number of particles in cell k by N_k. The total number of particles is N. The relative frequency of a particle to be found in cell k is then given by

$$\frac{N_k}{N} = p_k \ , \qquad k = 1, 2, \ldots, M \ . \tag{3.16}$$

p_k may be considered as the distribution function of the particles over the cells k. Because the cells have equal size and do not differ in their physical properties, we expect that the particles will be found with equal probability in each cell, i.e.,

$$p_k = \frac{1}{M} \ . \tag{3.17}$$

We now want to derive this result (3.17) from the properties of information. Indeed the information may be as follows: Before we make a measurement or obtain a message, there are R possibilities or, in other words, $K \ln R$ is a measure of our ignorance. Another way of looking at this is the following: R gives us the number of realizations which are possible in principle.

Now let us look at an ensemble of C containers, each with N gas atoms. We assume that in each container the particles are distributed according to different distribution functions p_k, i.e.,

$$p_k^{(1)}, p_k^{(2)}, p_k^{(3)}, \ldots \ .$$

Accordingly, we obtain different numbers of realizations, i.e., different information. For example, if $N_1 = N$, $N_2 = N_3 = \cdots = 0$, we have $p_1^{(1)} = 1$, $p_2^{(1)} = p_3^{(1)} = \cdots = 0$ and thus $I^{(1)} = 0$. On the other hand, if $N_1 = N_2 = N_3 = \cdots = N/M$, we have $p_1^{(2)} = 1/M$, $p_2^{(2)} = 1/M$, ..., so that $I^{(2)} = -M \log_2 (1/M) = M \log_2 M$, which is a very large number if the number of boxes is large.

Thus when we consider any container with gas atoms, the probability that it is one with the second distribution function is much greater than one with the first distribution function. That means there is an overwhelming probability of finding that probability distribution p_k realized which has the greatest number of possibilities

R and thus the greatest information. Hence we are led to require that

$$-\sum p_i \ln p_i = Extremum \tag{3.18}$$

under the constraint that the total sum of the probabilities p_i equals unity

$$\sum_{i=1}^{M} p_i = 1 \ . \tag{3.19}$$

This principle will turn out to be fundamental for application to realistic systems in physics, chemistry, and biology and we shall come back to it later.

The problem (3.18) with (3.19) can be solved using the method of Lagrange multipliers. This method consists in multiplying (3.19) by a still unknown parameter λ and adding it to the left-hand side of (3.18) now requiring that the total expression becomes an extremum. Here we are now allowed to vary the p_i's independently of each other, not taking into account the constraint (3.19). Varying the left-hand side of

$$-\sum p_i \ln p_i + \lambda \sum p_i = Extr \ . \tag{3.20}$$

means taking the derivative of it with respect to p_i which leads to

$$-\ln p_i - 1 + \lambda = 0 \ . \tag{3.21}$$

Equation (3.21) has the solution

$$p_i = e^{\lambda - 1} \tag{3.22}$$

which is independent of the index i, i.e., the p_i's are all equal. Inserting them into (3.19) we may readily determine λ so that

$$M e^{\lambda - 1} = 1 \ , \tag{3.23}$$

or, in other words, we find

$$p_i = \frac{1}{M} \tag{3.24}$$

in agreement with (3.17) as expected.

3.2 Information Gain

The expression (3.15) for the information can be interpreted as an average over f_j,

$$i = \sum p_j f_j \ , \quad \text{where} \tag{3.25}$$

$$f_j = -K \ln p_j \ , \qquad p_j \neq 0 \tag{3.26}$$

and the weight is p_j.