Preface

Complex Nonlinearity: Chaos, Phase Transitions, Topology Change and Path Integrals is a graduate–level monographic textbook. This is a book about *prediction & control* of general *nonlinear dynamics* of high–dimensional *complex systems* of various physical and non-physical nature and their underpinning geometro–topological change.

The book starts with a textbook–like expose on *nonlinear and chaotic dynamics* (Chapter 1). After an introduction into attractors and chaos, a brief history of chaos theory is given. Then, temporal chaotic dynamics is developed, both in its continuous form (nonlinear differential equations) and in its discrete form (nonlinear iteration maps). Spatio–temporal chaotic dynamics of nonlinear partial differential equations follows with some physiological examples. The Chapter ends with modern techniques of chaos–control, both temporal and spatio–temporal.

The *dynamical edge of chaos* physically corresponds to the *phase transitions*. Therefore, Chapter 2 continues exposé on complex nonlinearity, from the point of view of phase transitions and the related field of *synergetics*. After the introduction and classification of equilibrium phase transitions, a brief on Landau's theory is given (providing a background for order–parameters and synergetics). The concept is subsequently generalized into non–equilibrium phase transitions, together with important examples of oscillatory, fractal and noise–induced transitions. This core Chapter of the book also introduces the concept of partition function, together with its general, path–integral description. After that the basic elements of Haken's synergetics are presented, and subsequently developed into synergetics of *attractor neural networks*.

While the natural stage for linear dynamics comprises of flat, Euclidean geometry (with the corresponding calculation tools from linear algebra and analysis), the natural stage for nonlinear dynamics is curved, *Riemannian geometry* (with the corresponding tools from tensor algebra and analysis). In both cases, the system's (kinetic) energy is defined by the metric form, either Euclidean or Riemannian. The extreme nonlinearity – chaos – corresponds to the *topology change* of this curved geometrical stage, usually called configu-

ration manifold. Chapter 3 elaborates on geometry and topology change in relation with complex nonlinearity and chaos.

Chapter 4 develops general nonlinear dynamics, both continuous and discrete, deterministic and stochastic, in the unique form of *path integrals* and their *action–amplitude formalism*. This most natural framework for representing both phase transitions and topology change starts with *Feynman's sum over histories*, to be quickly generalized into the *sum over geometries and topologies*. This Chapter also gives a brief on general dynamics of fields and strings, as well as a path–integral based introduction on the chaos field theory. The Chapter concludes with a number of non–physical examples of complex nonlinear systems defined by path integrals.

The last Chapter puts all the previously developed techniques together and presents the *unified form of complex nonlinearity*. Here we have chaos, phase transitions, geometrical dynamics and topology change, all working together in the form of path integrals. The concluding section is devoted to discussion of hard vs. soft complexity, using the synergetic example of human bio-mechanics.

The objective of the present monograph is to provide a serious reader with a serious scientific tool that will enable them to actually *perform* a competitive research in modern complex nonlinearity. The monograph includes a comprehensive bibliography on the subject and a detailed index. For all mathematical questions, the reader is referred to our book *Applied Differential Geometry: A Modern Introduction*. World Scientific, Singapore, 2007.

Target readership for this monograph includes all researchers and students of complex nonlinear systems (in physics, mathematics, engineering, chemistry, biology, psychology, sociology, economics, medicine, etc.), working both in industry (i.e., clinics) and academia.

Adelaide,
Feb. 2008

V. Ivancevic, Defence Science & Technology Organisation, Australia, e-mail: Vladimir.Ivancevic@dsto.defence.gov.au

T. Ivancevic, School of Mathematics, The University of Adelaide, e-mail: Tijana.Ivancevic@adelaide.edu.au

2

Phase Transitions and Synergetics

The *dynamical edge of chaos* physically corresponds to the popular *phase transitions*. In this Chapter we present the basic principles of phase transitions, which are used in *synergetics*. After the introduction and classification of equilibrium phase transitions, a brief on Landau's theory is given (providing a background for order–parameters and synergetics). The concept is subsequently generalized into non–equilibrium phase transitions, together with important examples of oscillatory, fractal and noise–induced transitions. This core Chapter of the book also introduces the concept of partition function, together with its general, path–integral description. After that the basic elements of Haken's synergetics are presented, and subsequently developed into synergetics of *attractor neural networks.*

2.1 Introduction to Phase Transitions

Recall that in thermodynamics, a phase transition (or phase change) is the transformation of a thermodynamic system from one phase to another (see Figure 2.1). The distinguishing characteristic of a phase transition is an abrupt change in one or more physical properties, in particular the heat capacity, with a small change in a thermodynamic variable such as the temperature.

2.1.1 Equilibrium Phase Transitions

In *thermodynamics*, a *phase transition* represents the transformation of a system from one phase to another. Here the therm *phase* denotes a set of states of a macroscopic physical system that have relatively uniform chemical composition and physical properties (i.e., density, crystal structure, index of refraction, and so forth.) The most familiar examples of phases are solids, liquids, and gases. Less familiar phases include plasmas, Bose–Einstein condensates and fermionic condensates and the paramagnetic and ferromagnetic phases of magnetic materials.

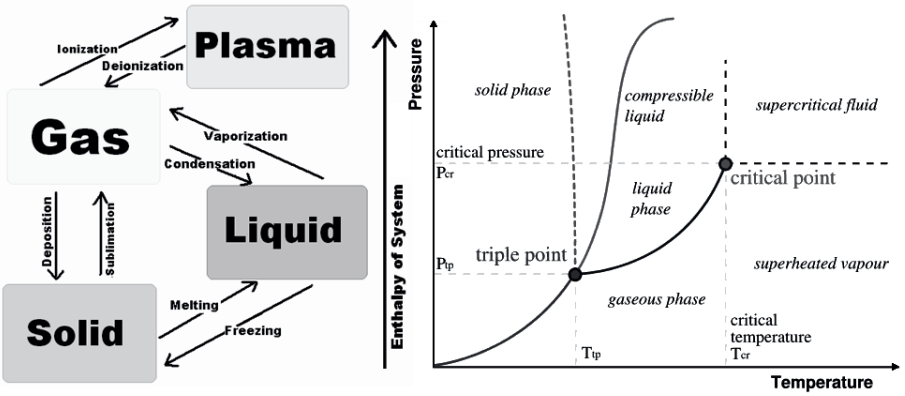

Fig. 2.1. Thermodynamic (or, equilibrium) phase transitions. Left: the term *phase transition* is most commonly used to describe transitions between solid, liquid and gaseous states of matter, in rare cases including plasma. Right: the transitions between the solid, liquid, and gaseous phases of a single component, due to the effects of temperature and/or pressure (adapted and modified from [Wik07].)

In essence, all thermodynamic properties of a system (entropy, heat capacity, magnetization, compressibility, and so forth) may be expressed in terms of the *free energy potential* $\mathcal{F}$ and its partial derivatives. For example, the *entropy* S is the first derivative of the free energy $\mathcal{F}$ with respect to the temperature T, i.e., $S = -\partial\mathcal{F}/\partial T$, while the *specific heat capacity* C is the second derivative, $C = T\,\partial S/\partial T$. As long as the free energy $\mathcal{F}$ remains analytic, all the thermodynamic properties will be well–behaved.

Now, the distinguishing characteristic of a phase transition is an *abrupt sudden change in one or more physical properties*, in particular the specific heat c, with a small change in a thermodynamic variable such as the temperature T. Standard examples of phase transitions are (see e.g., [LL78, Wik07]):

1. The transitions between the solid, liquid, and gaseous phases (boiling, melting, sublimation, etc.)
2. The transition between the ferromagnetic and paramagnetic phases of magnetic materials at the Curie point.
3. The emergence of superconductivity in certain metals when cooled below a critical temperature.
4. Quantum condensation of bosonic fluids, such as Bose–Einstein condensation and the superfluid transition in liquid helium.
5. The breaking of symmetries in the laws of physics during the early history of the universe as its temperature cooled.

When a system goes from one phase to another, there will generally be a stage where the free energy is non–analytic. This is known as a phase transition. Familiar examples of phase transitions are melting (solid to liquid),

freezing (liquid to solid), boiling (liquid to gas), and condensation (gas to liquid). Due to this *non–analyticity*, the free energies on either side of the transition are two different functions, so one or more thermodynamic properties will behave very differently after the transition. The property most commonly examined in this context is the heat capacity. During a transition, the heat capacity may become infinite, jump abruptly to a different value, or exhibit a 'kink' or *discontinuity* in its derivative[1].

Therefore, phase transitions come about when the free energy of a system is non–analytic for some choice of thermodynamic variables. This non–analyticity generally stems from the interactions of an extremely large number of particles in a system, and does not appear in systems that are too small.

2.1.2 Classification of Phase Transitions

Ehrenfest Classification

The first attempt at classifying phase transitions was the *Ehrenfest classification scheme*, which grouped phase transitions based on the degree of non-analyticity involved. Though useful, Ehrenfest's classification is flawed, as we will discuss in the next section.

Under this scheme, phase transitions were labelled by the lowest partial derivative of the free energy that is discontinuous at the transition. *First–order phase transitions* exhibit a discontinuity in the first derivative of the free energy with respect to a thermodynamic variable. The various solid⇒liquid⇒gas transitions are classified as first–order transitions, as the density, which is the first partial derivative of the free energy with respect to chemical potential, changes discontinuously across the transitions[2]. *Second–order phase transitions* have a discontinuity in a second derivative of the free energy. These include the ferromagnetic phase transition in materials such as iron, where the magnetization, which is the first derivative of the free energy with the applied magnetic field strength, increases continuously from zero as the temperature is lowered below the Curie temperature. The magnetic susceptibility, the second derivative of the free energy with the field, changes discontinuously. Under the *Ehrenfest classification scheme*, there could in principle be third, fourth, and higher–order phase transitions.

[1] In practice, each type of phase is distinguished by a handful of relevant thermodynamic properties. For example, the distinguishing feature of a solid is its rigidity; unlike a liquid or a gas, a solid does not easily change its shape. Liquids are distinct from gases because they have much lower compressibility: a gas in a large container fills the container, whereas a liquid forms a puddle in the bottom. Not all the properties of solids, liquids, and gases are distinct; for example, it is not useful to compare their magnetic properties. On the other hand, the ferromagnetic phase of a magnetic material is distinguished from the paramagnetic phase by the presence of bulk magnetization without an applied magnetic field.

[2] The pressure must be continuous across the phase boundary in equilibrium.

Modern Classification

The Ehrenfest scheme is an inaccurate method of classifying phase transitions, for it is based on the *mean–field theory* of phases, which is inaccurate in the vicinity of phase transitions, as it neglects the role of thermodynamic fluctuations. For instance, it predicts a finite discontinuity in the heat capacity at the ferromagnetic transition, which is implied by Ehrenfest's definition of second–order transitions. In real ferromagnets, the heat capacity diverges to infinity at the transition.

In the modern classification scheme, phase transitions are divided into two broad categories, named similarly to the Ehrenfest classes:

- The *first–order phase transitions*, or, *discontinuous phase transitions*, are those that involve a latent heat. During such a transition, a system either absorbs or releases a fixed (and typically large) amount of energy. Because energy cannot be instantaneously transferred between the system and its environment, first–order transitions are associated with *mixed–phase regimes* in which some parts of the system have completed the transition and others have not. This phenomenon is familiar to anyone who has boiled a pot of water: the water does not instantly turn into gas, but forms a turbulent mixture of water and water vapor bubbles. Mixed–phase systems are difficult to study, because their dynamics are violent and hard to control. However, many important phase transitions fall in this category, including the solid⇒liquid⇒gas transitions.
- The *second–order phase transitions* are the *continuous phase transitions*. These have no associated latent heat. Examples of second–order phase transitions are the ferromagnetic transition, the superfluid transition, and Bose–Einstein condensation.

2.1.3 Basic Properties of Phase Transitions

Critical Points

In systems containing liquid and gaseous phases, there exist a special combination of pressure and temperature, known as the *critical point*, at which the transition between liquid and gas becomes a second–order transition. Near the critical point, the fluid is sufficiently hot and compressed that the distinction between the liquid and gaseous phases is almost non–existent.

This is associated with the phenomenon of critical opalescence, a milky appearance of the liquid, due to density fluctuations at all possible wavelengths (including those of visible light).

Symmetry

Phase transitions often (but not always) take place between phases with different symmetry. Consider, for example, the transition between a fluid (i.e., liquid

or gas) and a crystalline solid. A fluid, which is composed of atoms arranged in a disordered but homogenous manner, possesses continuous translational symmetry: each point inside the fluid has the same properties as any other point. A crystalline solid, on the other hand, is made up of atoms arranged in a regular lattice. Each point in the solid is *not* similar to other points, unless those points are displaced by an amount equal to some lattice spacing.

Generally, we may speak of one phase in a phase transition as being more symmetrical than the other. The transition from the more symmetrical phase to the less symmetrical one is a *symmetry–breaking* process. In the fluid–solid transition, for example, we say that continuous translation symmetry is broken.

The ferromagnetic transition is another example of a symmetry–breaking transition, in this case the symmetry under reversal of the direction of electric currents and magnetic field lines. This symmetry is referred to as 'up–down symmetry' or 'time–reversal symmetry'. It is broken in the ferromagnetic phase due to the formation of magnetic domains containing aligned magnetic moments. Inside each domain, there is a magnetic field pointing in a fixed direction chosen spontaneously during the phase transition. The name *time–reversal symmetry* comes from the fact that electric currents reverse direction when the time coordinate is reversed.

The presence of symmetry–breaking (or nonbreaking) is important to the behavior of phase transitions. It was pointed out by Landau that, given any state of a system, one may unequivocally say whether or not it possesses a given symmetry [LL78]. Therefore, it cannot be possible to analytically deform a state in one phase into a phase possessing a different symmetry. This means, for example, that it is impossible for the solid–liquid phase boundary to end in a critical point like the liquid–gas boundary. However, symmetry–breaking transitions can still be either first or second order.

Typically, the more symmetrical phase is on the high–temperature side of a phase transition, and the less symmetrical phase on the low–temperature side. This is certainly the case for the solid–fluid and ferromagnetic transitions. This happens because the Hamiltonian of a system usually exhibits all the possible symmetries of the system, whereas the low–energy states lack some of these symmetries (this phenomenon is known as spontaneous symmetry breaking.) At low temperatures, the system tends to be confined to the low–energy states. At higher temperatures, thermal fluctuations allow the system to access states in a broader range of energy, and thus more of the symmetries of the Hamiltonian.

When symmetry is broken, one needs to introduce one or more extra variables to describe the state of the system. For example, in the ferromagnetic phase one must provide the net magnetization, whose direction was spontaneously chosen when the system cooled below the Curie point. Such variables are instances of *order parameters*. However, note that order parameters can also be defined for symmetry–nonbreaking transitions.

Symmetry–breaking phase transitions play an important role in cosmology. It has been speculated that, in the hot early universe, the vacuum (i.e., the various quantum fields that fill space) possessed a large number of symmetries. As the universe expanded and cooled, the vacuum underwent a series of symmetry–breaking phase transitions. For example, the electroweak transition broke the $SU(2) \times U(1)$ symmetry of the electroweak field into the $U(1)$ symmetry of the present–day electromagnetic field. This transition is important to understanding the asymmetry between the amount of matter and antimatter in the present–day universe.

Critical Exponents and Universality Classes

Continuous phase transitions are easier to study than first–order transitions due to the absence of latent heat, and they have been discovered to have many interesting properties. The phenomena associated with continuous phase transitions are called *critical phenomena*, due to their association with critical points.

It turns out that continuous phase transitions can be characterized by parameters known as critical exponents. For instance, let us examine the behavior of the heat capacity near such a transition. We vary the temperature T of the system while keeping all the other thermodynamic variables fixed, and find that the transition occurs at some critical temperature T_c. When T is near T_c, the heat capacity C typically has a *power law* behavior: $C \sim |T_c - T|^{-\alpha}$. Here, the constant α is the critical exponent associated with the heat capacity. It is not difficult to see that it must be less than 1 in order for the transition to have no latent heat. Its actual value depends on the type of phase transition we are considering. For -1 $< \alpha <$ 0, the heat capacity has a 'kink' at the transition temperature. This is the behavior of liquid helium at the 'lambda transition' from a normal state to the superfluid state, for which experiments have found $\alpha = -0.013 \pm 0.003$. For $0 < \alpha < 1$, the heat capacity diverges at the transition temperature (though, since $\alpha < 1$, the divergence is not strong enough to produce a latent heat.) An example of such behavior is the 3D ferromagnetic phase transition. In the $3D$ Ising model for uniaxial magnets, detailed theoretical studies have yielded the exponent $\alpha \sim 0.110$.

Some model systems do not obey this power law behavior. For example, mean–field theory predicts a finite discontinuity of the heat capacity at the transition temperature, and the 2D Ising model has a logarithmic divergence. However, these systems are an exception to the rule. Real phase transitions exhibit power law behavior.

Several other critical exponents – β, γ, δ, ν, and η – are defined, examining the power law behavior of a measurable physical quantity near the phase transition.

It is a remarkable fact that phase transitions arising in different systems often possess the same set of critical exponents. This phenomenon is known as *universality*. For example, the critical exponents at the liquid–gas critical point have been found to be independent of the chemical composition

of the fluid. More amazingly, they are an exact match for the critical exponents of the ferromagnetic phase transition in uniaxial magnets. Such systems are said to be in the same *universality class*. Universality is a prediction of the renormalization group theory of phase transitions, which states that the thermodynamic properties of a system near a phase transition depend only on a small number of features, such as dimensionality and symmetry, and is insensitive to the underlying microscopic properties of the system.

2.1.4 Landau's Theory of Phase Transitions

Landau's theory of phase transitions is a simple but powerful empirical thermodynamic theory by which the behavior of crystals at phase transitions can be described. It is based simply on a power series expansion of the free energy of the crystal with respect to one or a few prominent parameters distorting the symmetry of the crystal. The symmetry of the distortion decides which terms may be present, and which not. For example, odd terms on the power series expansion often are not allowed because the energy of the system is symmetric with respect to positive or negative distortion. With Landau's theory, the thermodynamics of the crystal (free energy, entropy, heat capacity) can be directly linked to it's structural state (volume, deviation from high symmetry, etc.), and both can be described as they change as a function of temperature or pressure.

More precisely, in Landau's theory, the probability (density) distribution function f is exponentially related to the potential $\mathcal{F}$,

$$f \approx \mathrm{e}^{-\mathcal{F}(T)}, \tag{2.1}$$

if $\mathcal{F}$ is considered as a function of the order parameter o. Therefore, the most probable order parameter is determined by the requirement $\mathcal{F} = \min$.

When $M_\uparrow$ elementary magnets point upwards and $M_\downarrow$ elementary magnets point downwards, the magnetization order parameter o is given by

$$o = (M_\uparrow - M_\downarrow)\, m, \tag{2.2}$$

where m is the magnetic moment of a single elementary magnet.

We expand the potential $\mathcal{F} = \mathcal{F}(o, T)$ into a power series of o,

$$\mathcal{F}(o,T) = \mathcal{F}(0,T) + \mathcal{F}'(0,T)o + \cdots + \frac{1}{4!}\mathcal{F}''''(0,T)o^4 + \cdots, \tag{2.3}$$

and discus $\mathcal{F}$ as a function of o. In a number of cases $\mathcal{F}' = \mathcal{F}''' = 0$, due to inversion symmetry. In this case, $\mathcal{F}$ has the form

$$\mathcal{F}(o,T) = \mathcal{F}(0,T) + \frac{\sigma}{2}o^2 + \frac{\beta}{4}o^4, \tag{2.4}$$

where $\beta > 0$, and $\sigma = a(T - T_c),\ (a > 0)$, i.e., it changes its sign at the critical temperature $T = T_c$.

Recall that the (negative) first partial derivative of the free energy potential $\mathcal{F}$ with respect to the *control parameter* – temperature T is the entropy

$$S = -\frac{\partial \mathcal{F}(q,T)}{\partial T}. \tag{2.5}$$

For $T > T_c$, $\sigma > 0$, and the minimum of $\mathcal{F}$ lies at $o = o_0 = 0$, and

$$S = S_0 = -\frac{\partial \mathcal{F}(0,T)}{\partial T}.$$

Also recall that the second partial derivative of $\mathcal{F}$ with respect to T is the specific heat capacity (besides the factor T)

$$C = T\frac{\partial S}{\partial T}. \tag{2.6}$$

One may readily check that S is continuous at $T = T_c$ for $\sigma = 0$. However, when we calculate the specific heat we get two different expressions above and below the critical temperature and thus a discontinuity at $T = T_c$.

Closely related to the Landau's theory of phase transitions is *Ginzburg–Landau model* of *superconductivity* (named after *Nobel Laureates Vitaly L. Ginzburg and Lev D. Landau*). It does not purport to explain the microscopic mechanisms giving rise to superconductivity. Instead, it examines the macroscopic properties of a superconductor with the aid of general thermodynamic arguments. Based on Landau's previously–established theory of second–order phase transitions, Landau and Ginzburg argued that the free energy F of a superconductor near the superconducting transition can be expressed in terms of a complex *order parameter* ψ, which describes how deep into the superconducting phase the system is. By minimizing the free energy with respect to fluctuations in the order parameter and the vector potential, one arrives at the *Ginzburg–Landau equation*, which is a generalization of the *nonlinear Schrödinger equation*.

2.1.5 Example: Order Parameters in Magnetite Phase Transition

The mechanism of the phase transition in magnetite (Fe_3O_4) at $T_V = 122$ K, discovered by Verwey [Ver39], has remained a big puzzle in the condensed matter physics for almost 70 years. Developments in experimental and theoretical methods during last years enabled to reveal subtle changes in the crystal and electronic structure below T_V [RAP01, LYA04]. A simple charge ordering picture in which metal–insulator transition is induced by electrostatic interactions was replaced by a highly complex scenario in which lattice, charge, spin and orbital degrees of freedom are involved. Recent theoretical studies revealed the important role of the local electron interactions and orbital correlations in the t_{2g} states on iron ions [LYA04].

The electronic interactions are complemented by the lattice deformation, which breaks the cubic symmetry and induces a low–temperature (LT) monoclinic phase driven by the electron–phonon interactions. In the work of [PPO06, PPO07], the phonon spectrum of magnetite has been analyzed using the *ab initio* computational technique [PLK97]. We have identified two primary *order parameters* (OPs) at $\mathbf{k}_X = \frac{2\pi}{a}(0,0,1)$ and $\mathbf{k}_\Delta = \frac{2\pi}{a}(0,0,\frac{1}{2})$ with the X_3 and Δ_5 symmetry, respectively, which both play important role in the Verwey transition (VT): (i) the Δ_5 mode is responsible for the doubling of the unit cell along the c direction in the monoclinic phase, while (ii) the X_3 phonon induces the *metal–insulator transition* by its coupling to the electronic states near the Fermi energy [PPO06]. Due to the *electron–phonon interaction* the above OPs are combinations of the electron (charge–orbital) and lattice components. This explains why the phonon soft mode has not been observed. Instead, low–energy critical fluctuations of OPs were found by the diffuse neutron scattering [FSY75]. The condensation of the OPs below T_V explains the crystal symmetry change as well as the charge-orbital ordering.

The group theory predicts also secondary OPs, which do not effect the symmetry below T_V but modify the properties of magnetite close to a transition point. At the Γ point, the T_{2g} mode can be classified as the secondary OP, and its coupling to the shear strain explains the softening of the C_{44} elastic constant [SNM05]. The lowest T_{2g} optic mode, marked in Fig. 1, could contribute quantitatively to the free energy, but it does not play any significant role for the VT.

In this section, following [PPO07], we analyze the *Landau free energy* for the VT, and discuss a solution corresponding to the LT monoclinic phase. We derive also the temperature dependence of C_{44}.

Free Energy

The Landau free energy can be expanded into a series of the components of the OPs. The invariant terms describing couplings between the OPs were derived using the group theory methods [SH02]. The only nonzero components of the primary OPs X_3 and Δ_5 are denoted by g and q, respectively. We include also the secondary OP with the T_{2g} symmetry (η) and shear–strain (ϵ). The free energy can be written in the form [PPO07]

$$\begin{aligned}\mathcal{F} = \mathcal{F}_0 &+ \frac{\alpha_1}{2}g^2 + \frac{\beta_1}{4}g^4 + \frac{\gamma_1}{6}g^6 + \frac{\alpha_2}{2}q^2 + \frac{\beta_2}{4}q^4 + \frac{\delta_1}{2}g^2q^2 \\ &+ \frac{\alpha_3}{2}\eta^2 + \frac{\alpha_4}{2}\epsilon^2 + \frac{\delta_2}{2}\eta g^2 + \frac{\delta_3}{2}\epsilon g^2 + \delta_4\eta\epsilon,\end{aligned} \tag{2.7}$$

were $\mathcal{F}_0$ is the part of the potential, which does not change through the transition. We assume that $\beta_1 > 0$, $\beta_2 > 0$ and $\gamma_1 > 0$ to ensure the stability of the potential at high temperatures. For the second–order terms we assume standard temperature behavior $\alpha_i = a_i(T - T_{ci})$ near the critical temperature T_{ci} for $i = 1, 2, 3$, which would correspond to a continuous phase transition.

The coefficient α_4 is the shear elastic constant at high temperatures (C_{44}^0). The coupling between the primary OPs is biquadratic, between the secondary and primary OPs has the linear–quadratic form, and the coupling between the components of the secondary OP is of the bilinear type. Taking first derivatives of $\mathcal{F}$ over all OPs we get [PPO07]

$$\begin{aligned}
\frac{\partial \mathcal{F}}{\partial g} &= g(\alpha_1 + \beta_1 g^2 + \gamma_1 g^4 + \delta_1 q^2 + \delta_2 \eta + \delta_3 \epsilon) = 0, \\
\frac{\partial \mathcal{F}}{\partial q} &= q(\alpha_2 + \beta_2 q^2 + \delta_1 g^2) = 0, \\
\frac{\partial \mathcal{F}}{\partial \eta} &= \alpha_3 \eta + \delta_4 \epsilon + \frac{\delta_2}{2} g^2 = 0, \\
\frac{\partial \mathcal{F}}{\partial \epsilon} &= \alpha_4 \epsilon + \delta_4 \eta + \frac{\delta_3}{2} g^2 = 0.
\end{aligned}$$

The solution $g = q = \eta = \epsilon = 0$ corresponds to the high–temperature cubic symmetry ($Fd\bar{3}m$). We obtain the dependence between g and q,

$$q^2 = -\frac{\delta_1 g^2 + \alpha_2}{\beta_2}, \tag{2.8}$$

which has three possible solutions: (i) $g = 0$ and $q^2 = -\frac{\alpha_2}{\beta_2}$ if $\alpha_2 < 0$ ($Pbcm$), (ii) $q = 0$ and $g^2 = -\frac{\alpha_2}{\delta_1}$ if $\alpha_2 > 0$ and $\delta_1 > 0$ or $\alpha_2 < 0$ and $\delta_1 > 0$ ($Pmna$), (iii) $g \neq 0$ and $q \neq 0$ ($P2/c$). In the brackets we put the space group symbols, which characterize the low–symmetry phases. The solution (iii) which corresponds to the experimentally observed LT monoclinic phase requires simultaneous condensation of both primary OPs. The necessary condition for this is a negative value of δ_1. Indeed, it has been established by the *ab initio* studies that the total energy is lowered when the crystal is distorted by both X_3 and Δ_5 modes [PPO07]. For $\delta_1 < 0$, (2.8) has a non–zero solution provided that $|\delta_1| g^2 > \alpha_2$. It implies that for $\alpha_2 > 0$ ($T > T_{c2}$), the phase transition occurs when the OP g exceeds a critical value $\frac{\alpha_2}{|\delta_1|}$, so it has a discontinuous (first–order) character.

From above relations we get [PPO07]

$$\eta = \frac{\delta_3 \delta_4 - \delta_2 \alpha_4}{2\alpha_3 \alpha_4 - 2\delta_4^2} g^2 \equiv \lambda_1 g^2, \qquad \epsilon = \frac{\delta_2 \delta_4 - \delta_3 \alpha_3}{2\alpha_3 \alpha_4 - 2\delta_4^2} g^2 \equiv \lambda_2 g^2,$$

which shows that $\eta \neq 0$ and $\epsilon \neq 0$ only if $g \neq 0$. Eliminating q, η and ϵ, the potential $\mathcal{F}$ can be written as a function of g

$$\mathcal{F} = \mathcal{F}_0' + \frac{\alpha}{2} g^2 + \frac{\beta}{4} g^4 + \frac{\gamma_1}{6} g^6,$$

where the renormalized coefficients are

$$\mathcal{F}_0' = \mathcal{F}_0 - \frac{\alpha_2^2}{4\beta_2}, \qquad \alpha = \alpha_1 - \frac{\alpha_2\delta_1}{\beta_2},$$
$$\beta = \beta_1 - \frac{\delta_1^2}{\beta_2} + 2\alpha_3\lambda_1^2 + 2\alpha_4\lambda_2^2 + 2\delta_2\lambda_1 + 2\delta_3\lambda_2 + 4\delta_4\lambda_1\lambda_2.$$

The zero–order and second–order terms depend on the parameters belonging to the primary OPs. The secondary OPs modify only the forth–order term. In this notation, the solution which minimizes the potential $\mathcal{F}$ reads

$$g_o^2 = \frac{-\beta + \sqrt{\beta^2 - 4\gamma\alpha}}{2\gamma}, \qquad q_o^2 = -\frac{\delta_1 g_o^2 + \alpha_2}{\beta_2},$$
$$\eta_o = \lambda_1 g_o^2, \qquad \epsilon_o = \lambda_2 g_o^2.$$

To study the softening of C_{44}, we have expressed the free energy as a function of ϵ only. In these calculations we have omitted the sixth–order term, which usually has a small contribution near the transition point. The elastic constant C_{44} is obtained using the standard definition

$$C_{44}(T) = \frac{\partial^2\mathcal{F}}{\partial\epsilon^2} = C_{44}^0 - \frac{\delta^2}{\alpha_3'} - \frac{\delta_3^2}{\beta_1'}, \qquad \text{where} \tag{2.9}$$

$$\delta = \delta_4 - \frac{\delta_2\delta_3}{2\beta_1'}, \qquad \alpha_3' = \alpha_3 - \frac{\delta_2^2}{2\beta_1'} = a_3(T - T_{c3}'), \qquad \beta_1' = \beta_1 - \frac{\delta_1^2}{\beta_2},$$

with $T_{c3}' = T_{c3} + \delta_2^2/2\beta_1'$. The second and third term in (2.9) are negative at high temperatures, so both contribute to the softening of C_{44}. It means that all couplings included in (2.7) are involved in this behavior. The main temperature dependence is caused by the second term, but also the last term in (2.9) may depend on temperature. Omitting the last term, (2.9) can be written in the form [PPO07]

$$C_{44} = C_{44}^0 \frac{T - T_0}{T - T_{c3}'}, \tag{2.10}$$

where $T_0 = T_{c3}' + \delta^2/C_{44}^0 a_3$.

For more technical details, see [PPO07] and references therein.

2.1.6 Universal Mandelbrot Set as a Phase–Transition Model

The problem of stability of time evolution is one of the most important in physics. Usually one can make the motion stable or unstable by changing some parameters which characterize Hamiltonian of the system. Stability regions can be represented on the phase diagram and transitions between them are described by catastrophe theory [Tho89, Arn78]. It can seem that a physical system or a mechanism can be taken from one domain of stability to any other by continuous and quasi–static variation of these parameters, i.e., that

the phase diagram is connected. However sometimes this expectation is wrong, because domains of stability can be separated by points where our system is getting totally destroyed. Unfortunately today it is too difficult to explore the full phase diagram for generic physical system with many parameters. Therefore, following [LL03], it was proposed in [DM06] to consider as a simpler model the discrete dynamics of one complex variable. The phase diagram in this case is known as Universal Mandelbrot Set (UMS). MS is a well–known object in mathematics (see [Wik07, Fer23], as well as Figure 1.44 above), but its theory is too formal and not well adjusted to the use in the phase transition theory. In this section, following [Mor07], we will make MS more practical for physical applications.

Structure of MS

First of all we remind the definition of MS and UMS from [DM06], which different from conventional definition in mathematical literature, see below. Mandelbrot Set (MS) is a set of points in the complex c plane. MS includes a point c if the map $x \to f(x,c)$ has stable periodic orbits. As shown in Figure2.2 MS consists of many clusters connected by trails, which in turn consist of smaller clusters and so on. Each cluster is linear connected and can be divided into elementary domains where only one periodic orbit is stable. Different elementary domains can merge and even overlap. Boundary of elementary domain of nth order, i.e., of a domain where an nth order orbit is stable, is a real curve $c(\alpha)$ given by the system:

$$\begin{cases} G_n(x,c) = 0 \\ F'_n(x,c) + 1 = e^{i\alpha} \end{cases}, \qquad \text{with}$$

$$F_n(x,c) = f^{\circ n}(x,c) - x,$$
$$G_n(x,c) = \frac{F_n(x,c)}{\prod_m G_m(x,c)}, \qquad (n = 1, ..., m).$$

Indeed, when $G_n(x,c)$ vanishes then x belongs to the orbit of exactly the nth order. This orbit is stable if $|\frac{\partial}{\partial x} f^{\circ n}(x,c)| < 1$, what implies the above equation. The solution of this system may give us more than a single nth order domain. Domains of different orders merge at the points c where

$$Resultant_x\left(G_n(x,c), G_k(x,c)\right) = 0. \tag{2.11}$$

Clearly, two orbits and thereafter domains merge only if n is divisor of k, so it is reasonable to consider only $Resultant_x(G_n, G_{mn}) = 0$, with $m = 2, 3 \ldots$ and $Discriminant_x(G_n)$ for $k = n$.

Physically MS is a phase diagram of discrete dynamic of one complex variable. It is clear from Figure 2.2 that one should distinguish between three types of connectivity in different places of phase diagram. The first is linear

connectivity: the possibility to connect any two points with a continuous line. The second type is weak connectivity: it means that only a closure of our set has linear connectivity. The third type we call strong connectivity: it means that any two interior points are connected with a thick tube.

Entire MS on Figure 2.2 is weakly[3] but not linearly connected and its clusters are linearly, but not strongly connected. Universal Mandelbrot Set (UMS) is unification of MS of different 1_c-parametric families. When we rise from MS to UMS we add more parameters to the base function. Thus entire UMS could become strongly connected, but it is unclear whether this really happens.

Fig. 2.2. The simplest examples of Mandelbrot sets $MS(x^2+c)$ and $MS(x^3+c)$, constructed by Fractal Explorer [Fer23]. The picture explains the terms 'clusters','elementary domain' and 'trails'. It is difficult to see any clusters except for the central one in the main figure, therefore one of the smaller clusters is shown in a separate enlarged picture to the left (adapted and modified from [Mor07]).

To explore MS of different functions we need to draw it. The method of MS construction which we can derive from the definition of MS is following. We are constructing the domains where different orbits are stable. We can build them using the fact that if orbit is stable then absolute value of derivative $\frac{df}{dx}$ is less than one.

Simplification of the Resultant Condition

In this section we prove that the resultant condition (2.11) can be substituted a by much simpler one [Mor07]:

$$Resultant_x\left(G_n(x),(F'_n+1-e^{i\alpha})\right)=0, \qquad \alpha=\frac{2\pi}{m}k.$$

To prove this equation, it is enough to find the points where

$$Resultant_x(F_n,\frac{F_{nm}(x)}{F_n(x)})=0.$$

[3] MS is usually claimed to be locally connected [Wik07], i.e., any arbitrary small vicinity of a point of MS contains a piece of some cluster. In our opinion weak connectivity is another feature, especially important for physical applications.

Then:

$$Resultant_x(F_n, \frac{F_{nm}(x)}{F_n(x)}) = Resultant_x(F_n, \frac{F'_{nm}(x)}{F'_n(x)}) = 0.$$

By definition of $F(x)$,

$$F_{nm}(x) = F_{n(m-1)}\left(F_n(x) + x\right) + F_n(x).$$

Then:

$$\frac{F'_{nm}(x)}{F'_n(x)} = \frac{F'_{n(m-1)}(x)}{F'_n(x)}(F'_n(x)+1)+1 \;=\; \frac{F'_{n(m-1)}(x)}{F'_n(x)}e^{i\alpha}+1 =$$
$$= (\ldots((e^{i\alpha}+1)e^{i\alpha}+1)e^{i\alpha}+1)\ldots)+1 = \sum_{l=0}^{m-1} e^{li\alpha} = \frac{e^{mi\alpha}-1}{e^{i\alpha}-1}.$$

Thus (2.11) implies that $e^{mi\alpha} - 1 = 0$ and therefore $\alpha = \frac{2\pi k}{m}$.

This theorem is a generalization of a well known fact for the central cardioid domain of $MS(x^2 + c)$ (see for example [Wik07]). This also provides a convenient parametrization of generic MS.

A Fast Method for MS simulation

Historically MS was introduced in a different way from our formulation. We call it $\widetilde{MS}$. It depends not only on the family of functions, but also on a point x_0. If c belongs to the $\widetilde{MS}(f, x_0)$ then

$$\lim_{n\to\infty} f^{\circ n}(x_0) \neq \infty$$

In the literature one usually puts $x_0 = 0$ independently of the shape of $f(x)$. Such $\widetilde{MS(f,0)} \neq MS(f)$, except for the families like $f = x^a + c$. Existing computer programs [Fer23] generate $\widetilde{MS}(f, 0)$, and can not be used to draw the proper $MS(f)$. Fortunately there is a simple relation [Mor07]:

$$MS(f) = \bigcup_{x_{cr}} \widetilde{MS}(f, x_{cr}).$$

where union is over all critical points of $f(x)$, $f'(x_{cr}) = 0$. Equation (2.1.6) is closely related to hyperbolic and local connectivity conjectures [Wik07]. It is also equivalent to the following two statements about the phase portrait in the complex x plane:

(I) If $\lim_{l\to\infty} f^{\circ l}(x_{cr}) \neq \infty$ then there is a stable periodic orbit O of finite order which attracts x_{cr}. It implies that

$$MS(f) \supseteq \bigcup_{x_{cr}} \widetilde{MS}(f, x_{cr}).$$

(II) If O is a stable periodic orbit, then a critical point x_{cr} exists, which is attracted to O. This implies that

$$MS(f) \subseteq \bigcup_{x_{cr}} \widetilde{MS}(f, x_{cr}).$$

The statement (I) says that if $\lim_{l\to\infty} f^{\circ l}(x_{cr}) \neq \infty$ then this limit exists and is a stable orbit with finite period, i.e., that there are no such things as strange attractors in discrete dynamics of one complex variable. This statement is unproved but we have no counter–examples.

The statement (II) is much easier. If x_0 is a stable fixed point, then it is surrounded by a disk–like domain, where $|f'(x)| < 1$. Its boundary is parametrized by $f'(x) = e^{i\alpha}$ and inside this area there is a point where $f'(x) = 0$, i.e., some critical point x_{cr} of f. It is important that this entire surrounding of x_0 – and thus this x_{cr} – lie inside the attraction domain of x_0:

$$|f(x_{cr}) - f(x)| < |x_{cr} - x|,$$

i.e., we found x_{cr} which is attracted to x_0. This argument can be easily extended to higher order orbits and can be used to prove (II). Equation (2.1.6) leads to a simple upgrade of programs, which construct MS.

Reducing $MS(x^3 + c)$ to $MS(x^2 + c)$

As application of our results we consider the 2_C-parametric section of UMS for the family

$$f(x) = a \cdot x^3 + (1 - a) \cdot x^2 + c,$$

which interpolates between $MS(x^3+c)$ at $a = 1$ and $MS(x^2+c)$ at $a = 0$. We extend consideration of [DM07a] to non–trivial second order clusters which were beyond the reach of the methods used in that paper.

For more details on MS, see [Mor07].

2.1.7 Oscillatory Phase Transition

Coherent oscillations are observed in neural systems such as the *visual cortex* and the *hippocampus*. The synchronization of the oscillators is considered to play important roles in *neural information processing* [GKE89]. There are mainly two viewpoints in the research of the oscillatory activity in neural systems. In the first viewpoint, the activity of each neuron is expressed by the firing rate, and the coherent oscillation appears owing to the interaction of the excitatory and inhibitory neurons. Wilson–Cowan and Amari found first oscillatory behavior theoretically in interacting neurons [WC72, Ama72]. Recently, [RRR02] proposed a more elaborate model to explain various EEG rhythms and epileptic seizures. If the spatial freedom is taken into consideration, the excitation wave can propagate. Wilson–Cowan performed numerical simulations of two layers of excitable neurons and inhibitory neurons [WC72].

In the second viewpoint, each neuron is regarded as an oscillator. Coherent oscillation appears as the global synchronization of the coupled oscillators. The global synchronization in general coupled oscillators was first studied by Winfree [Win67]. Kuramoto proposed a globally coupled phase oscillator model as a solvable model for the global synchronization [Kur84]. The leaky–integrate–fire model is one of the simplest models for a single neuron and often used to study dynamical behaviors of neural networks. Each neuron receives an input via synaptic connections from other neurons and it fires when the input goes over a threshold and sends out impulses to other neurons. In that sense, the coupling is instantaneous, and then the model is called *pulse–coupled oscillators*. Mirollo and Strogatz studied a globally coupled system of the *integrate–and–fire neurons*, and showed that perfect synchronization occurs in a finite time [MS90]. The synchronization of pulse coupled oscillators has been studied in deterministic systems by many researchers [TMS93, GR93, VAE94]. If each oscillator's behavior is stochastic, the model is generalized to a noisy phase oscillator model and a noisy integrate–and–fire model. In the stochastic system, the coherent oscillation appears as an analogue of the phase transition in the statistical mechanics. Globally coupled noisy phase oscillators were studied in [SK86, Kur91, Sak02, KS03], and globally coupled noisy integrate–and–fire model were studied in [BH99, Bru00, HNT01]. The globally coupled system is a useful model for the detailed analyzes, however, local or non–local interactions are more plausible, since neurons interact with other neurons via long axons or gap junctions. The non–locally coupled system of the deterministic integrate–and–fire neurons was also studied [GE01]. In this section, following [Sak04], we study a non–locally coupled noisy integrate–and–fire model with the direct numerical simulation of the *Fokker–Planck equation*.

The equation for a noisy integrate–and–fire neuron is written as

$$\dot{x} = 1 - bx + I_0 + \xi(t), \tag{2.12}$$

where x is a variable corresponding to the membrane potential, b is a positive parameter, I_0 denotes an external input, and $\xi(t)$ is the Gaussian white noise satisfying $\langle \xi(t)\xi(t')\rangle = 2D\delta(t-t')$. If x reaches a threshold 1, x jumps back to 0. If $b < 1 + I_0$, each neuron fires spontaneously. The Fokker–Planck equation for the Langevin equation (2.12) is [Sak04]

$$\partial_t P = -\frac{\partial}{\partial x}(1 - bx + I_0)P(x) + D\frac{\partial^2 P}{\partial x^2} + \delta(x)J_0(t), \tag{2.13}$$

where $J_0(t) = -D(\partial P/\partial x)_{x=1}$ is the firing rate. The stationary distribution $P_0(x)$ for the Fokker–Planck equation (2.13) is written as [Ric77]

$$\begin{aligned} P_0(x) &= P_0(0)\mathrm{e}^{\{(ax-(1/2)bx^2\}/D}, \qquad \text{for } x < 0, \\ &= P_0(0)\mathrm{e}^{\{(ax-(1/2)bx^2\}/D}\left[1 - \frac{\int_0^x \mathrm{e}^{\{-az+(1/2)bz^2\}/D}dz}{\int_0^1 \mathrm{e}^{\{-az+(1/2)bz^2\}/D}dz}\right], \quad \text{for } 0 < x < 1, \end{aligned} \tag{2.14}$$

where $a = 1 + I_0$ and $P(0)$ is determined from the normalization condition $\int_{-\infty}^{1} P_0(x)dx = 1$. The firing rate J_0 is determined as

$$J_0 = DP_0(0)/\int_0^1 \mathrm{e}^{\{-az+(1/2)bz^2\}/D}dz.$$

We have performed direct numerical simulation of (2.13) with the finite difference method with $\Delta x = 0.0002$ and $\Delta t = 2.5 \times 10^{-5}$, and checked that the stationary probability distribution (2.14) is successfully obtained.

We assume a non–locally coupled system composed of the noisy integrate–and–fire neurons. Each neuron interacts with other neurons via synaptic connections. Time delay exists generally for the synaptic connections. A model equation of the interacting noisy integrate–and–fire neurons is written as [Sak04]

$$\dot{x}_i = 1 - bx_i + I_i + \xi_i(t),$$

where x_i denotes the dimensionless membrane potential for the ith neuron, $\xi_i(t)$ denotes the noise term which is assumed to be mutually independent, i.e., $\langle \xi_i(t)\xi_j(t')\rangle = 2D\delta_{i,j}\delta(t-t')$, and I_i is the input to the ith neuron by the mutual interaction. The input I_i to the ith neuron from the other neurons is given by

$$I_i = \sum_j \sum_k g_{i,j} \frac{1}{\tau} \mathrm{e}^{-(t-t_k^j)/\tau}, \tag{2.15}$$

where t_k^j is the time of the kth firing for the jth neuron, $g_{i,j}$ denotes the interaction strength from the jth neuron to the ith neuron, and τ denotes a decay constant. The sum is taken only for $t > t_k^j$. The effect of the firing of the jth neuron to the ith neuron decays continuously with τ. If $\tau \to 0$, the coupling becomes instantaneous. Equation (2.15) is equivalent to

$$\dot{I}_i = -\{I_i - \sum_j \sum_k g_{i,j}\delta(t - t_k^j)\}/\tau.$$

If there are infinitely many neurons at each position y, we can define the number density of neurons with membrane potential x clearly at each position. The number density is expressed as $n(x, y, t)$ at position y and time t. The non–locally coupled system can be studied with a mean–field approach. In the mean–field approach, the number density is proportional to the probability distribution $P(x, y, t)$ for the probability variable x. The average value of $\delta(t - t_k^j)$ expresses the average firing rate at time t at the position y. It is expressed as $J_0(y, t) = -D(\partial n/\partial x)_{x=1}$. The number density $n(x, y, t)$ therefore obeys the Fokker–Planck type equation [Sak04]

$$\begin{aligned}
\partial_t n(x, y) &= -\frac{\partial}{\partial x}(1 - bx + I(y, t))n(x, y) + D\frac{\partial^2 n}{\partial x^2} + \delta(x)J_0(y, t), \\
\dot{I}(y, t) &= -\{I(y, t) - J(y, t)\}/\tau, \\
J(y, t) &= \int g(y, y')J_0(y', t)dy',
\end{aligned}$$

where $g(y, y')$ is the coupling strength from the neuron located at y' to the one at y, and $I(y), J_0(y)$ are respectively the input and the firing rate for the neuron at y.

We have assumed that the time delay for the signal to transmit between y' and y can be neglected and $g(y, y')$ depends only on the distance $|y - y'|$, i.e., $g(y, y') = g(|y - y'|)$. As two simple examples of the non-local coupling, we use

$$g_1(y, y') = c\exp(-\kappa|y - y'|) - d, \quad g_2(y, y') = c\exp(-\kappa|y - y'|) - d\exp(-\kappa'|y - y'|).$$

These forms of the coupling imply that the interaction is excitable locally, but the interaction strength decreases with the distance $|y - y'|$, and it becomes inhibitory when $|y-y'|$ is large. This *Mexican–hat coupling* was used in several neural models [Ama77], especially to study the competitive dynamics in neural systems. Although two layer models of excitatory neuron layer and inhibitory neuron layer may be more realistic, we consider the above simpler one–layer model. The inhibitory interaction approaches a constant value $-d$ for the coupling g_1, and 0 for the coupling g_2. The system size is assumed to be $L = 10$ as a simple example, and the periodic boundary conditions for the space variable y are imposed. We choose the damping constants κ and κ', as the exponential function decays to almost 0 for the distance $|y-y'| \sim L$. Therefore, the dynamical behaviors do not depend on the system size L qualitatively in the second model. But the dynamical behaviors depend on the system size L in the first model, because the range of the inhibitory interaction is infinite in the model.

There is a stationary and uniform solution $n(x, y, t) = n_0(x)$ and $I(y, t) = I_0$ in the non–locally coupled equation. The uniform solution satisfies

$$\begin{aligned} n_0(x) &= n_0(0)\mathrm{e}^{\{ax-(1/2)bx^2\}/D}, \quad \text{for } x < 0, \\ &= n_0(0)\mathrm{e}^{\{ax-(1/2)bx^2\}/D}\left[1 - \frac{\int_0^x \mathrm{e}^{\{-az+(1/2)bz^2\}/D}dz}{\int_0^1 \mathrm{e}^{\{-az+(1/2)bz^2\}/D}dz}\right], \quad \text{for } 0 < x < 1, \end{aligned}$$

where the parameter a is determined by the self-consistent condition

$$a = 1 - g_0 D(\partial n_0(x)/\partial x)_{x=1}, \qquad \text{where} \qquad g_0 = \int g(y, y')dy'.$$

To study the linear stability of the stationary and uniform solution, we consider small deviations $\delta n(x, y, t) = n(x, y, t) - n_0(x)$ and $\delta I(y, t) = I - I_0$ from the uniform solution. The small deviations can be expressed with the Fourier series as

$$\delta n(x, y, t) = \sum \delta n_k(x, t)\exp(iky) \qquad \text{and} \qquad \delta I(y, t) = \sum \delta I_k \exp(iky)$$

under the periodic boundary conditions, where $k = 2\pi m/L$. The perturbations δn_k and δI_k obey coupled linear equations [Sak04]

$$\frac{\partial \delta n_k(x,t)}{\partial t} = -\frac{\partial}{\partial x}\{(1-bx+I_0)\delta n_k(x,t) + \delta I_k(t) n_0(x)\} + D\frac{\partial^2 \delta n_k}{\partial x^2} + \delta(x)\delta J_0(t),$$
$$\frac{d\delta I_k(t)}{dt} = -\{\delta I_k(t) - g'\delta J_{0k}(t)\}/\tau, \qquad \text{where} \tag{2.16}$$
$$\delta J_{0k}(t) = -D(\partial n_k/\partial x)_{x=1} and g' = \int g(y,y')\mathrm{e}^{ik(y'-y)}dy'.$$

For L is sufficiently large, $g' = 2c\kappa/(\kappa^2+k^2) - dL\delta_{k,0}$ for the coupling g_1 and $g' = 2c\kappa/(\kappa^2+k^2) - 2d\kappa'/(\kappa'^2+k^2)$ for the coupling g_2. The stability of the stationary state is determined by the real part of the eigenvalues of the linear equation (10). But, the stationary solution $n_0(x)$ is a nontrivial function of x, and it is not so easy to obtain the eigenvalues. Here we have evaluated the real part of the largest eigenvalue of the linear equation for various k by direct numerical simulations of (2.16). The dynamical behavior in the long time evolution of (2.16) is approximately expressed with the largest eigenvalue λ, that is, δn_k and $\delta I_k \sim \mathrm{e}^{\lambda t}$ for $t \gg 1$. We have numerically calculated the linear growth rate of the norm $\{\int(\delta n_k)^2 dx + (\delta I_k)^2\}^{1/2}$ (which grows as $\mathrm{e}^{(\mathrm{Re}\lambda)t}$ for $t \gg 1$) every time–interval 0.001. Since the norm grows to infinity or decays to zero in the natural time evolution of the linear equation, we have renormalized the variables every time–interval 0.001, as the norm is 1 by the rescaling $c\delta n_k \to \delta n_k$ and $c\delta I_k \to \delta I_k$ with a constant c.

Since the pulse propagates one round L with period $T = 3.23$, the velocity of the travelling pulse is $L/T \sim 3.1$. A regular limit cycle oscillation with period T is observed at each point. The directions depend on the initial conditions. The travelling pulse state is an ordered state in the non-locally coupled system. The locally excitable interaction facilitates the local synchronization of the firing, but the global inhibition suppresses the complete synchronization. As a result of the frustration, a travelling pulse appears. The pulse state is different form the travelling pulse observed in an excitable system, since the uniform state is unstable in our system and the pulse state is spontaneously generated from the stationary asynchronous state. The input $I(y,t)$ to the neuron at position y exhibits regular limit cycle oscillation.

As a second example, we consider a non–locally coupled system with the coupling function

$$g_2(y) = 1.8\exp(-4|y|) - 0.48\exp(-|y|).$$

A supercritical phase transition occurs at $D \sim 0.0155$, which is also consistent with the linear stability analysis. Near the critical value, the amplitude of the oscillation is small and the wavy state seems to be sinusoidal. As D is decreased, the oscillation amplitude increases and the sinusoidal waves change into pulse trains gradually. Pulses are created periodically near $x \sim 6$ and they are propagating alternatively in different directions. The inversely–propagating pulses collide at $x \sim 1$ and they disappear. Namely, there are a pacemaker region (a source region) and a sink region of travelling pulses in

this solution. This type of wavy state including a pacemaker region and the simple pulse-train state are bistable.

In summary, we have studied the non–locally noisy integrate–and–fire model with the Fokker–Planck equation. We have found that a travelling pulse appears as a result of oscillatory phase transitions. We found also a pulse-train state by changing the form of the interaction. The wavy states appear as a phase transition from a asynchronous state when the noise strength is decreased. We have investigated a 1D system for the sake of simplicity of numerical simulations, but we can generalize the model equation to a 2D system easily. Our non–locally coupled integrate–and–fire model might be too simple, however, the wavy state is one of the typical dissipative structures far from equilibrium. Therefore, the spontaneously generated waves might be observed as some kind of brain waves also in real neural systems. For more details, see [Sak04].

2.1.8 Partition Function and Its Path–Integral Description

Recall that in *statistical mechanics*, the *partition function* Z is used for statistical description of a system in thermodynamic equilibrium. Z depends on the physical system under consideration and is a function of temperature T as well as other parameters (such as volume V enclosing a gas etc.). The partition function forms the basis for most calculations in statistical mechanics. It is most easily formulated in *quantum statistical mechanics* (see e.g., [Fey72]).

Classical Partition Function

A system subdivided into N subsystems, where each subsystem (e.g., a particle) can attain any of the energies $\epsilon_j \ \ (j = 1, ..., N)$, has the partition function given by the sum of its Boltzmann factors,

$$\zeta = \sum_{j=0}^{\infty} \mathrm{e}^{-\beta\epsilon_j},$$

where $\beta = \frac{1}{k_B T}$ and k_B is *Boltzmann constant*. The interpretation of ζ is that the probability that the subsystem will have energy ϵ_j is $\mathrm{e}^{-\beta\epsilon_j}/\zeta$. When the number of energies ϵ_j is definite (e.g., particles with spin in a crystal lattice under an external magnetic field), then the indefinite sum is replaced with a definite sum. However, the total partition function for the system containing N subsystems is of the form

$$Z = \prod_{j=1}^{N} \zeta_j = \zeta_1 \zeta_2 \zeta_3 \cdot ...,$$

where ζ_j is the partition function for the jth subsystem. Another approach is to sum over all system's total energy states,

$$Z = \sum_{r=1}^{N} \mathrm{e}^{-\beta E_r}, \qquad \text{where} \qquad E_j = n_1^{(j)}\epsilon_1 + n_2^{(j)}\epsilon_2 + ...$$

In case of a system containing N non–interacting subsystems (e.g., a real gas), the system's partition function is given by

$$Z = \frac{1}{N!}\zeta^N.$$

This equation also has the more general form

$$Z = \frac{1}{N!h^{3N}} \int \prod_{i=1}^{N} d^3q^i d^3p_i \sum_{i=1}^{N} \mathrm{e}^{-\beta H_i},$$

where $H_i = H_i(q^i, p_i)$ is the ith subsystem's Hamiltonian, while h^{3N} is a normalization factor.

Given the partition function Z, the system's *free energy* F is defined as

$$F = -k_B T \ln Z,$$

while the *average energy* U is given by

$$U = \frac{1}{Z} E_i \mathrm{e}^{-\frac{E_i}{k_B T}} = -\frac{d}{d\beta} (\ln Z) .$$

Liner Harmonic Oscillators in Thermal Equilibrium. The partition function Z, free energy F, and average energy U of the system of M oscillators can be found as follows: The oscillators do not interact with each other, but only with the *heat bath*. Since each oscillator is independent, one can find F_i of the ith oscillator and then $F = \sum_{i=1}^{M} F_i$. For each ith oscillator (that can be in one of N states) we have [Fey72]

$$Z_i = \sum_{n=1}^{N} \mathrm{e}^{-\frac{E_n^i}{k_B T}}, \qquad F_i = -k_B T \ln Z_i, \qquad U_i = \frac{1}{Z} \sum_{n=1}^{N} E_n^i \mathrm{e}^{-\frac{E_n^i}{k_B T}}.$$

Quantum Partition Function

Partition function Z of a quantum–mechanical system may be written as a trace over all states (which may be carried out in any basis, as the trace is basis–independent),

$$Z = \mathrm{Tr}(\mathrm{e}^{-\beta \hat{H}}),$$

where $\hat{H}$ is the system's Hamiltonian operator. If $\hat{H}$ contains a dependence on a parameter λ, as in $\hat{H} = \hat{H}_0 + \lambda \hat{A}$, then the statistical average over $\hat{A}$ may be found from the dependence of the partition function on the parameter, by differentiation,

$$< \hat{A} >= -\beta^{-1} \frac{d}{d\lambda} \ln Z(\beta, \lambda).$$

However, if one is interested in the average of an operator that does not appear in the Hamiltonian, one often adds it artificially to the Hamiltonian, calculates Z as a function of the extra new parameter and sets the parameter equal to zero after differentiation.

More general, in *quantum field theory*, we have a *generating functional* J of the *field* $\phi(q)$ and the partition function is usually expressed by the *Feynman path integral* [Fey72] (see Chapter 4)

$$Z[J] = \int \mathcal{D}[\phi] \mathrm{e}^{i(S[\phi] + \int d^N q\, J(q)\phi(q))},$$

where $S = S[\phi]$ is the *field action functional*.

Vibrations of Coupled Oscillators

In this subsection, following [Fey72], we give both classical and quantum analysis of vibrations of coupled oscillators. R. Feynman used this method as a generic model for the crystal lattice.

Consider a *crystal lattice* with A atoms per unit cell, such that $3A$ coordinates α must be given to locate each atom. Also let $Q_{\alpha,N}$ denote the *displacement* from equilibrium of the coordinate α in the Nth cell. $Q_{\alpha,N+M}$ is the displacement of an atom in a cell close to N.

The *kinetic energy* of the lattice is given by

$$T = \frac{1}{2} Q_{\alpha,N} Q_{\alpha,N},$$

while its *potential energy* (in linear approximation) is given by

$$V = \frac{1}{2} C^M_{\alpha\beta} Q_{\alpha,N} Q_{\beta,N+M}.$$

Classical Problem

The so–called *original Hamiltonian* is given by

$$H = \sum_i \frac{p_i\prime^2}{2m_i} + \frac{1}{2} C_{ij}\prime q^i\prime q^j\prime,$$

where the $q^i\prime$ are the coordinates of the amount of the lattice displacement from its equilibrium, $p_i\prime = m_i \dot{q}^i\prime$ are the canonical momenta, and $C_{ij}\prime = C_{ji}\prime$ are constants. To eliminate the mass constants m_i, let

$$q^i = q^i\prime \sqrt{m_i} \qquad \text{and} \qquad C_{ij} = \frac{C_{ij}\prime}{\sqrt{m_i m_j}}.$$

Then

$$p_i = \frac{\partial L}{\partial \dot{q}^i} = \frac{p_i\prime}{\sqrt{m_i}}, \qquad (L \text{ is the Lagrangian of the system})$$

and we get the *simplified Hamiltonian*

$$H = \frac{1}{2}\sum_i p_i^2 + \frac{1}{2}C_{ij}q^i q^j.$$

The Hamilton's equations of motion now read

$$\dot{q}^i = \partial_{p_i} H = p_i, \qquad \dot{p}_i = -\partial_{q^i} H = -C_{ij}q^j = \ddot{q}^i.$$

We now break the motion of the system into *modes*, each of which has its *own frequency* ω. *The total motion of the system is a sum of the motions of the modes.* Let the αth mode have frequency ω_α so that

$$q^i_{(\alpha)} = \mathrm{e}^{-\mathrm{i}\omega_\alpha t} a_i^{(\alpha)}$$

for the motion of the αth mode, with $a_i^{(\alpha)}$ independent of time. Then

$$\omega_\alpha^2 a_i^{(\alpha)} = C_{ij} a_j^{(\alpha)}.$$

In this way, the classical *problem of vibrations of coupled oscillators* has been reduced to the *problem of finding eigenvalues* and *eigenvectors* of the real, symmetric matrix $\|C_{ij}\|$. In order to get the ω_α we must solve the *characteristic equation*

$$\det \left\| C_{ij} - \omega^2 \delta_{ij} \right\| = 0.$$

Then the eigenvectors $a_i^{(\alpha)}$ can be found. It is possible to choose the $a_i^{(\alpha)}$ so that

$$a_i^{(\alpha)} a_i^{(\beta)} = \delta_{\alpha\beta}.$$

The general solution for q^i is

$$q^i = C_\alpha q^i_{(\alpha)},$$

where the C_α are arbitrary constants. If we take

$$Q_\alpha = C_\alpha \mathrm{e}^{-\mathrm{i}\omega_\alpha t},$$

we get

$$q^i = a_i^{(\alpha)} Q_\alpha.$$

From this it follows that

$$a_i^{(j)} q^i = a_i^{(j)} a_i^{(\alpha)} Q_\alpha = \delta_{\alpha j} Q_\alpha = Q_j.$$

Making the change of variables, $Q_j = a_i^{(j)} q^i$, we get $H = \sum_\alpha H_\alpha$, where

$$H_\alpha = \frac{1}{2}p_\alpha^2 + \frac{1}{2}\omega_\alpha^2 Q_\alpha.$$

This has the expected solutions: $Q_\alpha = C_\alpha \mathrm{e}^{-\mathrm{i}\omega_\alpha t}$.

Quantum–Mechanical Problem

Again we have the original Hamiltonian

$$H = \sum_i \frac{p_i\prime 2}{2m_i} + \frac{1}{2} C'_{ij} q^i\prime q^j\prime,$$

where this time

$$p_i\prime = \frac{1}{\mathrm{i}} \frac{\partial}{\partial q^i\prime} \qquad \text{(in normal units } \hbar = 1\text{)}.$$

Making the same change of variables as before, we get

$$Q_\alpha = a_i^{(\alpha)} q^i = a_i^{(\alpha)} \sqrt{m_i} q^i\prime,$$
$$H = \sum_\alpha H_\alpha, \qquad \text{where} \qquad H_\alpha = -\frac{1}{2} \frac{\partial^2}{\partial Q_\alpha^2} + \frac{1}{2} \omega_\alpha^2 Q_\alpha.$$

It follows immediately that the eigenvalues of our original Hamiltonian are

$$E = \sum_\alpha (N_\alpha + \frac{1}{2}) \omega_\alpha.$$

The solution of a quantum–mechanical system of coupled oscillators is trivial once we have solved the characteristic equation

$$0 = \det \left\| C_{ij} - \omega^2 \delta_{ij} \right\| = \det \left\| \frac{C_{ij\prime}}{\sqrt{m_i m_j}} - \omega^2 \delta_{ij} \right\|.$$

If we have a solid with $\frac{1}{3}(10^{23})$ atoms we must apparently find the eigenvalues of a 10^{23} by 10^{23} matrix. But if the solid is *crystal*, the problem is enormously simplified. The *classical Hamiltonian for a crystal* is

$$H = \frac{1}{2} \sum_{\alpha,N} \dot{Q}_{\alpha,N}^2 + \frac{1}{2} \sum_{\alpha,\beta,N,M} C_{\alpha\beta}^M Q_{\alpha,N} Q_{\beta,N+M},$$

and the *classical equation of motion for a crystal lattice* is (using $C_{\alpha\beta}^M = C_{\beta\alpha}^{-M}$)

$$\ddot{Q}_{\alpha,N} = -\sum_{M,\beta} C_{\alpha\beta}^M Q_{\beta,N+M}.$$

In a given mode, if one cell of the crystal is vibrating in a certain manner, it is reasonable to expect all cells to vibrate the same way, but with different phases. So we try

$$Q_{\alpha,N} = a_\alpha(K) \mathrm{e}^{-\mathrm{i}\omega t} \mathrm{e}^{\mathrm{i}K\cdot N},$$

where K expresses the relative phase between cells. The $\mathrm{e}^{\mathrm{i}K\cdot N}$ factor allows for wave motion. We now want to find the dispersion relations, or $\omega = \omega(K)$.

$$\omega^2 a_\alpha e^{iK\cdot N} = \sum_{M,\beta} (C^M_{\alpha\beta} a_\beta e^{iK\cdot M}) e^{iK\cdot N}.$$

Let

$$\gamma_{\alpha\beta}(K) = \sum_M C^M_{\alpha\beta} e^{iK\cdot M}$$

(note that $\gamma_{\alpha\beta}(K)$ is Hermitian).

Then $\omega^2 a_\alpha = \sum_\beta \gamma_{\alpha\beta} a_\beta$, and we must solve the characteristic equation of a $3A-$by$-3A$ matrix:

$$\det|\gamma_{\alpha\beta} - \omega^2 \delta_{\alpha\beta}| = 0.$$

The solutions of the characteristic equation are

$$\omega^{(r)}(K) = \omega(^r_K),$$

where r runs from 1 to $3A$. The motion of a particular mode can be written

$$Q^{(r)}_{\alpha,N}(K) = a^r_\alpha(K) e^{-i\omega^{(r)}(K)^r} e^{iK\cdot N},$$

where

$$a^r_\alpha a^{*r'}_\alpha = \delta_{rr'}.$$

Then the general motion can be described by

$$Q_{\alpha,N} = \sum_{K,r} C_r(K) a^r_\alpha(K) e^{-i\omega^{(r)}(K)^r} e^{iK\cdot N},$$

where $C_r(K)$ are arbitrary constants.

Let $Q_r(K) = C_r(K) e^{-i\omega^{(r)}(K)^r}$. $Q_r(K)$ describe the motion of a particular mode. Then we have

$$Q_{\alpha,N} = \sum_{K,r} Q_r(K) a^r_\alpha(K) e^{iK\cdot N}.$$

It follows that

$$Q_r(K) \propto \sum_{\alpha,N} Q_{\alpha,N} a^{*r}_\alpha(K) e^{-iK\cdot N}, \qquad (\propto \text{ means 'proportional to'}),$$

and the Hamiltonian for the system is

$$\begin{aligned} H &= \frac{1}{2} \sum_{\alpha,N} \left[\dot{Q}^2_{\alpha,N} + \sum_{\beta,M} C^M_{\alpha\beta} Q_{\alpha,N} Q_{\beta,N+M} \right] \\ &= \frac{1}{2} \sum_{K,r} \left[|\dot{Q}_r(K)|^2 + \omega^{2(r)}(K) |Q_r(K)|^2 \right]. \end{aligned}$$

A Cubic Lattice of Harmonic Oscillators

Assume the unit cell to be a cubic lattice with one atom per cell. Each atom behaves as an harmonic oscillator, with spring constants k_A (nearest neighbors), and k_B (diagonal–, or next–nearest neighbors). This case is fairly simple, and we can simplify the notation: $\alpha = 1, 2, 3$.

$$Q_{1,N} = X_N, \qquad Q_{2,N} = Y_N, \qquad Q_{3,N} = Q_N.$$

We wish to find the 3 natural frequencies associated with each k of the crystal. To do this, we must find $C^M_{\alpha\beta}$ and then $\gamma_{\alpha\beta}$. In complex coordinates,

$$V = \sum_{\alpha,\beta} V_{\alpha\beta}, \qquad \text{where} \qquad V_{\alpha\beta} = \sum_{N,M} C^M_{\alpha\beta} Q^*_{\alpha,N} Q_{\beta,N+M},$$

where * denotes complex conjugation. For example,

$$V_{11} = \sum_{N,M} C^M_{11} X^*_N X_{N+M}.$$

If we express the displacement of atom N from its normal position as X_N, then the *potential energy* from the distortion of the spring between atoms N and M is

$$\frac{1}{2} k_M \left[(X_N - X_{N+M}) \cdot \frac{M}{|M|} \right]^2,$$

where $k_M = k_A$ for $N + M$ a nearest neighbor to N, $k_M = k_B$ for $N + M$ a next–nearest neighbor.

In summing over N and M to get the total potential energy we must *divide V by two*, for we count each spring twice. If we use complex coordinates, however, we *multiply V by two* to get the correct equations of motion:

$$\begin{aligned}
V &= \frac{1}{2} \sum_{N,M} k_M \left[(X_N - X_{N+M}) \cdot \frac{M}{|M|} \right]^2, \\
V_{11} &= \frac{1}{2} \sum_{N,M} k_M \left(\frac{M_X}{|M|} \right)^2 (X^*_N - X^*_{N+M})(X_N - X_{N+M}) \\
&= \frac{1}{2} \sum_{N,M} k_M \left(\frac{M_X}{|M|} \right)^2 [(X^*_N X_N + X^*_{N+M} X_{N+M}) - (X^*_N X_{N+M} + X_N X^*_{N+M})] \\
&= \sum_{N,M} k_M \left(\frac{M_X}{|M|} \right)^2 [X^*_N X_N - X_N X^*_{N+M}] .
\end{aligned}$$

Comparing the above expressions we see that

$$C^0_{11} = 2k_A + 4k_B, \qquad C^{\pm(1,0,0)}_{11} = -k_A, \qquad \text{and so on.}$$

In this way, all the $C^M_{\alpha\beta}$ can be found. We can then calculate

$$\gamma_{\alpha\beta}(K) = \sum_M C^M_{\alpha\beta} \mathrm{e}^{\mathrm{i}K\cdot M}.$$

We wish to solve

$$\det|\gamma_{\alpha\beta} - \omega^2\delta_{\alpha\beta}| = 0.$$

For each relative phase K, there are 3 solutions for ω. Thus we get $3N$ values of ω and $\omega^{(r)}(K)$.

2.1.9 Noise–Induced Non–Equilibrium Phase Transitions

Noise is usually thought of as a phenomenon which perturbs the observation and creates disorder (see section 4.6.1). This idea is based mainly on our day to day experience and, in the context of physical theories, on the study of equilibrium systems. The effect of noise can, however, be quite different in *nonlinear non–equilibrium systems*. Several situations have been documented in the literature, in which the noise actually participates in the creation of ordered states or is responsible for surprising phenomena through its interaction with the nonlinearities of the system [HL84]. Recently, a quite spectacular phenomenon was discovered in a specific model of a spatially distributed system with multiplicative noise, white in space and time. It was found that the noise generates an *ordered symmetry–breaking state* through a genuine *second–order phase transition*, whereas no such transition is observed in the absence of noise [BPT94, BPT97].

Recently it has been shown that a white and *Gaussian multiplicative noise* can lead an *extended* dynamical system (fulfilling appropriate conditions) to undergo a *phase* transition towards an *ordered* state, characterized by a nonzero order parameter and by the breakdown of ergodicity [BPT94]. This result–first got within a Curie–Weiss–like *mean–field approximation*, and further extended to consider the simplest correlation function approach–has been confirmed through extensive numerical simulations [BPT97]. In addition to its *critical* nature as a function of the noise intensity σ, the newly found noise–induced phase transition has the noteworthy feature of being *reentrant*: for each value of D above a threshold one, the ordered state exists only inside a window $[\sigma_1, \sigma_2]$. At variance with the known case of *equilibrium* order⇒disorder transitions that are induced (in the simplest lattice models) by the nearest–neighbor coupling constant D and rely on the bi–stability of the local potential, the transition in the case at hand is led by the *combined effects* of D and σ through the nonlinearities of the system. Neither the zero–dimensional system (corresponding to the $D = 0$ limit) nor the deterministic one ($\sigma = 0$) show any transition.

General Zero–Dimensional System

To smoothly introduce the subject, we will start from the well–known *logistic equation*, and add to it a multiplicative white noise.

Noisy Logistic Equation

Recall that the logistic equation (also called the *Verhulst model* or *logistic growth curve*) is a model of population growth first published by P. Verhulst in 1845 (see [Wei05]). The model is continuous in time, but a modification of the continuous equation to a discrete quadratic recurrence equation known as the *logistic map* is widely used in *chaos theory*. The standard logistic equation

$$\dot{x} = \lambda x - x^2, \tag{2.17}$$

where the parameter λ is usually constrained to be positive, has a solution

$$x(t) = \frac{1}{1 + \left(\frac{1}{x_0} - 1\right) \mathrm{e}^{-\lambda t}}.$$

Now, if we add a *multiplicative zero–mean Gaussian white noise* $\xi = \xi(t)$ with *noise intensity* σ to (2.17), we get the *Langevin SDE* (stochastic differential equation)

$$\dot{x} = \lambda x - x^2 + x\,\xi. \tag{2.18}$$

If we apply the *Stratonovitch interpretation* to the Langevin equation (2.18), we get the corresponding *Fokker–Planck equation*

$$\partial_t P(x,t) = -\partial_t \left(\lambda x - x^2\right) P(x,t) + \frac{\sigma^2}{2} \partial_x x\, \partial_x P(x,t) \tag{2.19}$$

derermining the *probability density* $P(x,t)$ for the variable $x(t)$. The equation (2.19) has the *stationary probability density*

$$P_{st}(x) = \frac{1}{Z} x^{\frac{2\lambda}{\sigma^2} - 1} \exp\left(-\frac{2x}{\sigma^2}\right)$$

(where Z is a normalization constant), with *two extrema*:

$$x_1 = 0, \qquad x_2 = \lambda - \frac{\sigma^2}{2}.$$

General Zero–Dimensional Model

Now, following [BPT94, BPT97], we consider the following SDE that generalizes noisy logistic equation (2.18),

$$\dot{x} = f(x) + g(x)\,\xi, \tag{2.20}$$

where, as above, $\xi = \xi(t)$ denotes the Gaussian white noise with first two moments

$$\langle \xi(t) \rangle = 0, \qquad \langle \xi(t)\, \xi(t') \rangle = \sigma^2 \delta(t - t').$$

If we interpret equation (2.20) according to the Stratonovitch interpretation, we get the corresponding Fokker–Planck equation

$$\partial_t P(x,t) = -\partial_x \left[f(x) + P(x,t) \right] + \frac{\sigma^2}{2} \partial_x \left(g(x)\, \partial_x \left[g(x)\, P(x,t) \right] \right),$$

with the *steady–state solution*

$$P_{st}(x) = \frac{1}{Z} \exp \left(\int_0^x \frac{f(y) - \frac{\sigma^2}{2} g(y) g'(y)}{\frac{\sigma^2}{2} g^2(y)} dy \right), \tag{2.21}$$

where $g'(x)$ stands for the derivative of $g(x)$ with respect to its argument. The extrema of the steady–state probability density obey the following equation

$$f(x) - \frac{\sigma^2}{2} g(x) g'(x) = 0. \tag{2.22}$$

Note that this equation is not identical to the equation $f(x) = 0$ for the steady states in the absence of multiplicative noise. As a result, the most probable states need not coincide with the deterministic stationary states. More importantly, solutions can appear or existing solutions can be destabilized by the noise. These changes in the asymptotic behavior of the system have been generally named noise–induced phase transitions [HL84].

To illustrate this phenomenon, consider the case of a deterministically stable steady state at $x = 0$, e.g.,

$$f(x) = -x + o(x),$$

perturbed by a multiplicative noise. As is clear from equations (2.6–2.22), a noise term of the form

$$g(x) = 1 + x^2 + o(x^2)$$

will have a stabilizing effect, since

$$-(\sigma^2/2) g(x) g'(x) = -\sigma^2 x + o(x^2),$$

and it makes the coefficient of x more negative. On the other hand, noise of the form

$$g(x) = 1 - x^2 + o(x^2)$$

i.e., with maximal amplitude at the reference state $x = 0$, has the tendency to 'destabilize' the reference state. In fact, above a critical intensity $\sigma^2 > \sigma_c^2 = 1$, the stationary probability density will no longer have a maximum at $x = 0$, and 'noise–induced' maxima can appear. This phenomenon remains possible even if the deterministic steady–state equation, got by fixing the random value

of the noise to a constant value λ, namely, $f(x) + \lambda g(x) = 0$, has a unique solution for all λ [BPT94, BPT97].

Following the formalism for equilibrium states, it is tempting to introduce the notion of a 'stochastic potential' $U_{st}(x)$ by writing:

$$P_{st}(x) \sim \exp\left[-U_{st}(x)\right].$$

One concludes that for a system undergoing a noise–induced transition, e.g., for $g(x) = 1 - x^2 + o(x^2)$, and for $\sigma^2 > \sigma_c^2$, the stochastic potential has two minima. Consider now a spatially extended system got by coupling such units. The coupling is such that it favors the nearest–neighbor units, to stay at the same maximum of the probability density (minimum of the stochastic potential). In analogy to what happens for equilibrium models (such as the Landau–Ginzburg model), one expects that this system will undergo a phase transition for some critical value of the 'temperature' (noise intensity) σ^2. However, it turns out that this is not the case. It was shown in [BPT94, BPT97] that one needs a noise of precisely the other type, namely $g(x) = 1 + x^2 + o(x^2)$, to generate a genuine phase transition.

General $d-$Dimensional System

The general model has been introduced in [BPT94, BPT97]: a dD extended system of typical linear size L is restricted to a hypercubic lattice of $N = L^d$ points, whereas time is still regarded as a continuous variable. The state of the system at time t is given by the set of stochastic variables $\{x_i(t)\}$ $(i = 1, \ldots, N)$ defined at the sites $\mathbf{r}_i$ of this lattice, which obey a system of coupled ordinary SDEs (with implemented Stratonovich interpretation)

$$\dot{x}_i = f(x_i) + g(x_i)\,\eta_i + \frac{D}{2d} \sum_{j \in n(i)} (x_j - x_i), \tag{2.23}$$

Equations (2.23) are the discrete version of the *partial* SDE which in the continuum would determine the state of the extended system: we recognize in the first two terms the generalization of Langevin's equation for site i to the case of multiplicative noise (η_i is the *colored multiplicative noise* acting on site $\mathbf{r}_i$). For the specific example, perhaps the simplest one exhibiting the transition under analysis,

$$f(x) = -x(1 + x^2)^2, \qquad g(x) = 1 + x^2. \tag{2.24}$$

The last term in (2.23) is the lattice version of the Laplacian $\nabla^2 x$ of the extended stochastic variable $x(\mathbf{r}, t)$ in a reaction–diffusion scheme. $n(i)$ stands for the set of $2d$ sites which form the immediate neighborhood of the site $\mathbf{r}_i$, and the coupling constant D between neighboring lattice sites is the diffusion coefficient.

Here we want to investigate the effects of the self–correlation time τ of the multiplicative noise on the model system just described [MDT00]. To that end we must assume a specific form for the noises $\{\eta_i = \eta_i(t)\}$: we choose *Ornstein–Uhlenbeck noise*, i.e., Gaussian distributed stochastic variables with zero mean and exponentially decaying correlations,

$$\langle \eta_i(t)\, \eta_j(t') \rangle = \delta_{ij} (\sigma^2/2\tau) \exp(-|t - t'|/\tau). \tag{2.25}$$

They arise as solutions of an *un–coupled* set of Langevin SDEs,

$$\tau \dot{\eta}_i = -\eta_i + \sigma \xi_i \tag{2.26}$$

where the $\{\xi_i = \xi_i(t)\}$ are white noises–namely, Gaussian stochastic variables with zero mean and $\delta-$correlated:

$$\langle \xi_i(t)\, \xi_j(t') \rangle = \delta_{ij} \delta(t - t').$$

For $\tau \to 0$, the Ornstein–Uhlenbeck noise $\eta_i(t)$ approaches the white–noise limit $\xi_i^W(t)$ with correlations

$$\langle \xi_i^W(t)\, \xi_j^W(t') \rangle = \sigma^2 \delta_{ij} \delta(t - t').$$

Mean–Field Approximation

The mean–field approximation here follows closely Curie–Weiss' mean–field approach to magnetism (see [MDT00]), and consists in replacing the last term in (2.23)

$$\Delta_i \equiv \frac{D}{2d} \sum_{j \in n(i)} (x_j - x_i), \tag{2.27}$$

by

$$\bar{\Delta}_i \equiv D(\bar{x} - x_i), \tag{2.28}$$

where $\bar{x}$ is the *order parameter* that will be determined self–consistently. In other words, the (short–ranged) interactions are substituted by a time– and space–independent 'external' field whose value *depends on the state* of the system. Since in this approximation equations (2.23) get immediately decoupled, there is no use in keeping the subindex i and we may refer to the systems in (2.23) and (2.26) as if they were single equations (Hereafter, the primes will indicate derivatives with respect to x (clearly $\bar{\Delta}' = -D$)).

If we take the time derivative of (2.23), replace first $\dot{\eta}$ in terms of η and ξ from (2.26) and then η in terms of $\dot{x}$ and x from (2.23), we get the following *non–Markovian* SDE:

$$\tau(\ddot{x} - \frac{g'}{g}\dot{x}^2) = -\left(1 - \tau\left[(f + \bar{\Delta})' - \frac{g'}{g}(f + \bar{\Delta})\right]\right)\dot{x} + (f + \bar{\Delta}) + \sigma g \xi. \tag{2.29}$$

Now, following [MDT00], we perform an *adiabatic elimination* of variables, namely, neglecting $\ddot{x}$ and $\dot{x}^2$, so that the system's dynamics becomes governed

by a Fokker–Planck equation. The resulting equation, being *linear* in $\dot{x}$ (but not in x), can be immediately solved for $\dot{x}$, giving

$$\dot{x} = Q(x;\bar{x}) + S(x;\bar{x})\xi, \tag{2.30}$$

with

$$Q(x;\bar{x}) \equiv (f + \bar{\Delta})\theta, \tag{2.31}$$
$$S(x;\bar{x}) \equiv \sigma g \theta, \tag{2.32}$$
$$\theta(x;\bar{x}) \equiv \{1 - \tau g[(f + \bar{\Delta})/g]'\}^{-1}. \tag{2.33}$$

The Fokker–Planck equation associated to the SDE (2.30) is

$$\partial_t P(x,t;\bar{x}) = -\partial_x \left[R_1(x;\bar{x})P(x,t;\bar{x})\right] + \frac{1}{2}\partial_x^2 \left[R_2(x;\bar{x})P(x,t;\bar{x})\right], \tag{2.34}$$

with *drift* and *diffusion* coefficients given by

$$R_1(x;\bar{x}) = Q + \frac{1}{4}(S^2)' \tag{2.35}$$
$$R_2(x;\bar{x}) = S^2. \tag{2.36}$$

The solution of the time–independent Fokker–Planck equation leads to the stationary probability density

$$P_{st}(x;\bar{x}) = \frac{1}{Z}\exp\left[\int_0^x dx' \frac{2R_1(x';\bar{x}) - \partial_{x'}R_2(x';\bar{x})}{R_2(x';\bar{x})}\right]. \tag{2.37}$$

The value of $\bar{x}$ arises from a *self–consistency relation*, once we equate it to the average value of the random variable x_i in the stationary state

$$\bar{x} = \langle x \rangle \equiv \int_{-\infty}^{\infty} dx\, x\, P_{st}(x;\bar{x}) \equiv F(\bar{x}). \tag{2.38}$$

Now, the condition

$$\left.\frac{dF}{d\bar{x}}\right|_{\bar{x}=0} = 1 \tag{2.39}$$

allows us to find the *transition line* between the *ordered* and the *disordered phases*.

Results

The mean–field approximation of the general dD extended system are the following (see [MDT00]):

A. As in the white–noise case $\tau = 0$, the ordering phase transition is *reentrant with respect to* σ: for a range of values of D that depends on τ, ordered states can only exist within a window $[\sigma_1, \sigma_2]$. The fact that this window shifts to the right *for small* τ means that, for fixed D, color *destroys* order just above σ_1 but *creates* it just above σ_2.

B. For fixed $\sigma > 1$ and $\tau \neq 0$, ordered states exist *only within a window* of values for D. Thus the ordering phase transition is *also reentrant with respect to* D. For τ small enough the maximum value of D compatible with the ordered phase increases rather steeply with σ, reaching a maximum around $\sigma \sim 5$ and then decreases gently. For $\tau \geq 0.1$ it becomes evident (in the ranges of D and σ analyzed) that the region sustaining the ordered phase is *closed*, and shrinks to a point for a value slightly larger than $\tau = 0.123$.
C. For fixed values of $\sigma > 1$ and D larger than its minimum for $\tau = 0$, the system *always* becomes disordered for τ large enough. The maximum value of τ consistent with order altogether corresponds to $\sigma \sim 5$ and $D \sim 32$. In other words, ordering is possible *only* if the multiplicative noise inducing it has short memory.
D. The fact that the region sustaining the ordered phase finally shrinks to a point means that even for that small region in the $\sigma-$D plane for which order is induced by color, a further increase in τ destroys it. In other words, the phase transition is *also reentrant with respect to* τ. For D large enough there may exist even *two* such windows.

Order Parameter

As already mentioned above, the *order parameter* in this system is $m \equiv |\bar{x}|$, namely, the positive solution of the consistency equation (2.38). Consistently with what has been discussed in (A) and (C), we see that as τ increases the window of σ values where ordering occurs shrinks until it disappears. One also notices that at least for this D, the value of σ corresponding to the maximum order parameter varies very little with τ.

The *short–time evolution* of $\langle x \rangle$ can be obtained multiplying (2.34) by x and integrating:

$$\frac{d\langle x \rangle}{dt} = \int_{-\infty}^{\infty} dx R_1(x;\bar{x}) P(x,t;\bar{x}). \tag{2.40}$$

Let us assume an initial condition such that at early times $P(x, t \sim 0; \bar{x}) = \delta(x - \bar{x})$. Equating $\bar{x} = \langle x \rangle$ as before, we get the *order parameter equation*

$$\frac{d\langle x \rangle}{dt} = R_1(\bar{x}, \bar{x}). \tag{2.41}$$

The solution of (2.41) has an initial *rising* period (it is initially *unstable*) reaching very soon a maximum and tending to zero afterwards.

For $D/\sigma^2 \to \infty$, equation (2.41) is valid also in the *asymptotic regime* since $P_{st}(x) = \delta(x-\bar{x})$ [BPT97]. According to this criterion, in the $D/\sigma^2 \to \infty$ limit the system undergoes a second–order phase transition *if* the corresponding zero-dimensional model presents *a linear instability in its short–time dynamics*, i.e., if after linearizing (2.41):

$$\langle \dot{x} \rangle = -\alpha \langle x \rangle, \tag{2.42}$$

one finds that $\alpha < 0$. We then see that the trivial (disordered) solution $\langle x \rangle = 0$ is stable only for $\alpha > 0$. For $\alpha < 0$ other stable solutions with $\langle x \rangle \neq 0$ appear, and the system develops order through a genuine *phase* transition. In this case, $\langle x \rangle$ can be regarded as the *order parameter*. In the white noise limit $\tau = 0$ this is known to be the case for sufficiently large values of the coupling D and for a window of values for the noise amplitude $\sigma \in [\sigma_1, \sigma_2]$.

In summary, we have:

A. Multiplicative noise can shift or induce phase transitions in 0D systems.
B. Multiplicative noise can induce phase transitions in spatially extended systems.
C. Mean–field approximation predicts a minimal coupling strength for the appearance of noise induced phase transitions.
D. Mean–field approximation predicts, that the phase transition is reentrant, i.e., the ordered phase is destroyed by even higher noise intensity.
E. Appearance of an ordered phase results from a nontrivial cooperative effect between multiplicative noise, nonlinearity and diffusion.

2.1.10 Noise–Driven Ferromagnetic Phase Transition

Over the last decade, the dynamics of ferromagnetic systems below their critical temperatures in a periodically oscillating magnetic field have been studied both theoretically [TO90, LP90, Ach97, SRM98, KWR01, FTR01, YTY02] and experimentally [JY95]. The systems exhibit two qualitatively different behaviors referred to as *symmetry–restoring oscillation* (SRO) and *symmetry–breaking oscillation* (SBO), depending on the frequency Ω and the amplitude h of the applied magnetic field. It has been established that there exists a sharp transition line between SRO and SBO on the (Ω, h) plane, which is called the *dynamical phase transition* (DPT). The DPT was first observed numerically in the deterministic mean–field system for a ferromagnet in a periodically oscillating field [TO90], and has subsequently been studied in numerous *Monte Carlo simulations* of the kinetic Ising system below critical temperature [LP90, Ach97, SRM98, KWR01]. It has also been observed experimentally in [JY95].

Recently, the DPT was investigated by introducing the model equation [FTR01]

$$\dot{s}(t) = (T_c - T)s - s^3 + h\cos\Omega t.$$

This equation is a simplified model for the Ising spin system at the temperature T below its critical value T_c in an external periodic magnetic field. By appropriately scaling the magnetization s, time t, and the applied field, this equation is written as

$$\dot{s}(t) = s - s^3 + h\cos\Omega t. \tag{2.43}$$

The SBO and SRO are observed in (2.43) and the transition line between them on the (Ω, h) plane is determined analytically [FTR01].

It is quite interesting to ask whether DPT is observed under another kind of applied field, especially random field with bounded amplitude. The fundamental aim of the present section is to study the dynamics of $s(t)$ with a dichotomous Markov noise (DMN) $F(t)$ instead of periodically oscillating external field $h\cos\Omega t$ (see, e.g., [Kam92]).

The equation of motion [OHF06]

$$\dot{s} = f(s) + F(t), \tag{2.44}$$

with a nonlinear function $f(s)$ and the DMN $F(t)$ has been extensively studied by many authors (see, e.g., [KHL79, BBK02, HR83]). It is well known that the master equation for the system can be derived, and then transition phenomena of stationary probability densities concerning the intensity of $F(t)$, for example, are studied, which are referred to as the noise-induced phase transition [KHL79, HL84]. The asymptotic drift velocity $\langle\dot{s}\rangle$ in the case of $f(s)$ being periodic functions are also discussed as a specific dynamic property [BBK02]. Furthermore, the *mean first–passage time* (MFPT) and transition rates are investigated as another important dynamic property when $f(s)$ is the force associated with the bistable potential given by (2.44).

Symmetry–Breaking Phase Transition

Model Equation and Noise–Induced Phase Transition

We consider the equation of motion driven by the external field $F(t)$ [OHF06],

$$\frac{ds(t)}{dt} = f(s) + F(t), \qquad (f(s) = s - s^3) \tag{2.45}$$

where $F(t)$ is a symmetric DMN with taking the values $\pm H_0$. Here the probability $p(\tau)$ that $F(t)$ continues to take the identical value $+H_0$ or $-H_0$ longer than time τ is given by

$$p(\tau) = \mathrm{e}^{-\tau/\tau_f}. \tag{2.46}$$

This implies that the correlation time of $F(t)$ is equal to $\tau_f/2$. Throughout this section, numerical integrations of (2.45) are carried out by using the Euler difference scheme with the time increment $\Delta t = 1/100$.

Without DMN, $s(t)$ eventually approaches either of the stationary fixed points ± 1, one of which is achieved according to the initial condition $s(0)$ as shown in Figure 2.3. In the presence of DMN, if $H_0 < H_c$, H_c being defined by

$$H_c \equiv 2(1/3)^{3/2} = 0.3849\cdots, \tag{2.47}$$

then $f(s) + H_0 = 0 (f(s) - H_0 = 0)$ has three real roots $s_{j+}(s_{j-})$, ($j = 1$, 2, and 3). Each value of $s_{j\pm}$ is graphically shown in Figure 2.3(a). On the other hand, if $H_0 > H_c$, then $f(s) + H_0 = 0$ ($f(s) - H_0 = 0$) has only one real root s_+ (s_-) given by

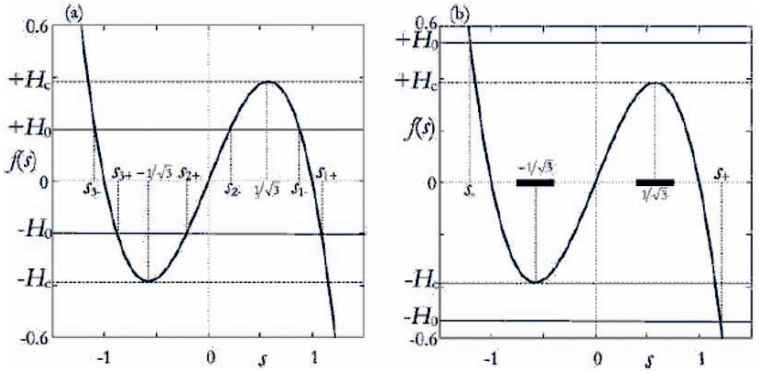

Fig. 2.3. The function $f(s) = s - s^3$ is shown by solid lines. (a) Real roots $s_{j\pm}$, $(j = 1, 2, 3)$ for $H_0 < H_c$ and (b) real roots $s_\pm$ for $H_0 > H_c$ of the algebraic equation $s - s^3 \pm H_0 = 0$. The definition of H_c is graphically represented (adapted and modified from [OHF06]).

$$s_\pm = \left[\frac{1}{2}\left(\pm H_0 + \sqrt{H_0^2 - H_c^2}\right)\right]^{1/3} + \left[\frac{1}{2}\left(\pm H_0 - \sqrt{H_0^2 - H_c^2}\right)\right]^{1/3}, \quad (2.48)$$

which are indicated in Figure 2.3(b). Next let us consider the dynamics described by (2.45) for $H_0 < H_c$ and for $H_0 > H_c$, and discuss similarity and difference between the dynamics in the periodically oscillating field case and those in the present DMN case. A part of our results belongs to the context of the noise–induced phase transition and MFPT in [KHL79, HL84, BBK02, HR83]. In the case of $H_0 < H_c$, three motions numerically integrated are shown in Figures 2.4(a) and (b). Two motions confined in the ranges $s_{1-} < s(t) < s_{1+}$ and $s_{3-} < s(t) < s_{3+}$ are both stable. The long time average $\langle s(t)\rangle$ of each motion does not vanish, and the motion is called SBM in relation to DPT in the oscillating external field case. On the other hand, the motion $s_u(t)$ confined in the range $s_{2+} < s_u(t) < s_{2-}$ is unstable. The long time average of $s_u(t)$ vanishes, and in this sense the motion is called SRM. It should be noted that this unstable SRM is located between two stable SBM, which has a similar characteristic to SBO of DPT [FTR01].

The motion of $s(t)$ for $H_0 > H_c$ is shown in Figures 2.4(c) and (d). One observes that there exists a stable SRM confined in the range $s_- < s(t) < s_+$. For SRM, the time average of $s(t)$ vanishes, i.e., $\langle s(t)\rangle = 0$. The comparison between Figures 2.4(b) and (d) suggests that the SRM for $H_0 > H_c$ is generated via the "attractor merging crisis" [Ott93] of the two SBM's and one unstable SRM, i.e., the two SBM's and one unstable SRM disappear and then one stable SRM takes place at $H_0 = H_c$. This situation is similar to that in the DPT case. However, in contrast to the DPT case, the transition line on the (τ_f^{-1}, H_0) plane is independent of the correlation time τ_f of $F(t)$ and the average $\langle s(t)\rangle$ depends discontinuously on H_0.

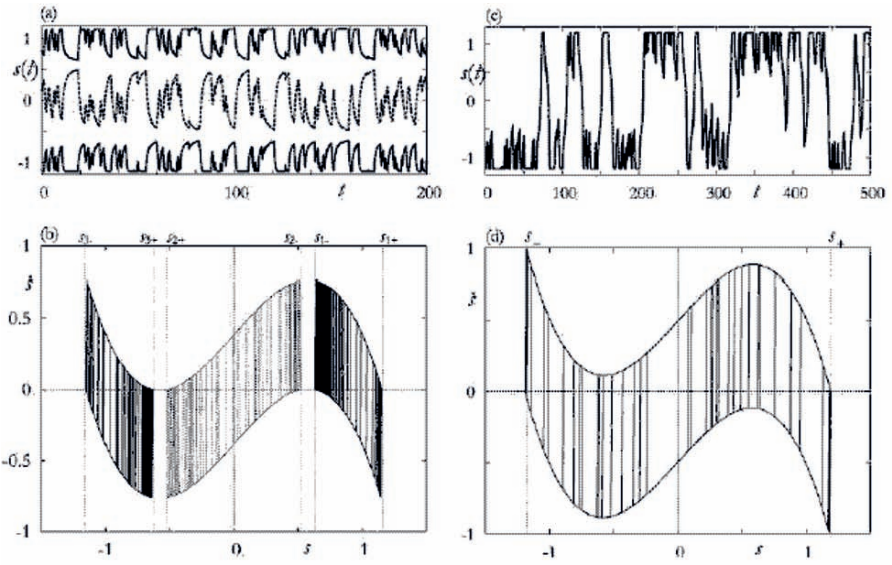

Fig. 2.4. Figures (a) and (b) show the motions obtained by numerically integrating (2.45) for $H_0 = 0.38(< H_c)$ and $\tau_f = 5$, where two SBM's (solid line) and an unstable SRM (dashed line) are drawn. The unstable SRM is evaluated by replacing $t \to -t$. On the other hand, Figures (c) and (d) show the motions for $H_0 = 0.5(> H_c)$ and $\tau_f = 10$ (adapted and modified from [OHF06]).

Stationary Distribution Functions

In this subsection, we discuss the stationary distribution functions for SBM and SRM. To this aim, we first consider a slightly general nonlinear Langevin equation of motion driven by DMN [OHF06],

$$\dot{x}(t) = f(x) + g(x)F(t), \tag{2.49}$$

where $f(x)$ and $g(x)$ are generally nonlinear functions of x and $F(t)$ is DMN [Ris84]. The temporal evolution of the distribution function $P(x, F, t)$ that $x(t)$ and $F(t)$ respectively take the values x and $F(= \pm H_0)$ is determined by [KHL79, HL84]

$$\begin{aligned} \frac{\partial}{\partial t}P(x,t) &= -\frac{\partial}{\partial x}\left[f(x)P(x,t) + H_0 g(x)q(x,t)\right], \\ \frac{\partial}{\partial t}q(x,t) &= -\frac{2}{\tau_f}q(x,t) - \frac{\partial}{\partial x}\left[f(x)q(x,t) + H_0 g(x)P(x,t)\right], \end{aligned} \tag{2.50}$$

where we put $P(x,t) \equiv P(x,+H_0,t)+P(x,-H_0,t)$ and $q(x,t) \equiv P(x,+H_0,t) - P(x,-H_0,t)$. The stationary distribution $P^{st}(x) \equiv P(x,\infty)$ is solved to yield

$$P^{st}(x) = \tag{2.51}$$
$$N\frac{g(x)}{H_0^2 g(x)^2 - f(x)^2}\exp\left\{-\frac{1}{\tau_f}\int^x dx'\left[\frac{1}{f(x') - H_0 g(x')} + \frac{1}{f(x') + H_0 g(x')}\right]\right\},$$

provided that each of the equations

$$\dot{x} = f(x) + H_0 g(x), \qquad \dot{x} = f(x) - H_0 g(x)$$

has at least one stable fixed point, where N is the normalization constant.

By substituting $f(x) = x - x^3$ and $g(x) = 1$, ((2.45)), into (2.51), the stationary distribution function $P^{st}_{SBM}(s)$ for SBM ($H_0 < H_c$) for $s_{3-} < s < s_{3+}$ or $s_{1-} < s < s_{1+}$ is written as

$$P^{st}_{SBM}(s) \propto |s^2 - s^2_{1+}|^{-\beta_{1+}}|s^2 - s^2_{1-}|^{-\beta_{1-}}|s^2 - s^2_{2+}|^{-\beta_{2+}}, \tag{2.52}$$
$$\beta_{j\pm} = 1 - \tau_f^{-1}|(s_{j\pm} - s_{k\pm})(s_{j\pm} - s_{l\pm})|,$$

where $(j,k,l) = (1,2,3)$, (2,3,1), and (3,1,2).

On the other hand, the stationary distribution function $P^{st}_{SRM}(s)$ for the SRM ($H_0 > H_c$) for $s_- < s < s_+$ is obtained as

$$P^{st}_{SRM}(s) \propto |s^2 - s^2_+|^{\frac{\tau_f^{-1}}{3s_+^2-1}-1}\left[(s^2 + s^2_+ - 1)^2 - s^2_+ s^2\right]^{-\frac{\tau_f^{-1}}{3s_+^2-1}-1}$$
$$\times \exp\left\{\frac{\tau_f^{-1}s_+}{(s^2_+ - 1/3)\sqrt{3s^2_+ - 4}}\left[\arctan\left(\frac{2s - s_+}{\sqrt{3s^2_+ - 4}}\right)\right.\right.$$
$$\left.\left. - \arctan\left(\frac{2s + s_+}{\sqrt{3s^2_+ - 4}}\right)\right]\right\}. \tag{2.53}$$

As H_0 is increased, the form of the stationary distribution function changes drastically from the forms in (2.52) to (2.53) at $H_0 = H_c$. This phenomenon which is induced by the disappearance of two pairs of stable and unstable fixed points [BBK02] is an example of the *noise–induced phase transitions* [HL84].

It turns out that the transition line between SRM and SBM on the (τ_f^{-1}, H_0) plane is given by $H_0 = H_c$. Furthermore, the long time average of $s(t)$, $\langle s(t)\rangle$, depends discontinuously on H_0 at $H_0 = H_c$. These behaviors are quite different from those of the DPT case driven by periodically oscillating field, $F(t) = h\cos(\Omega t)$ [Ach97, FTR01]. The transition point h_c for a fixed Ω between SRO and SBO depends on the frequency Ω, and $\langle s(t)\rangle$ is a continuous function of h.

MFPT Through the Channels

We hereafter discuss the dynamics for H_0 slightly above H_c. Let us first consider the behavior obeying the equations [OHF06]

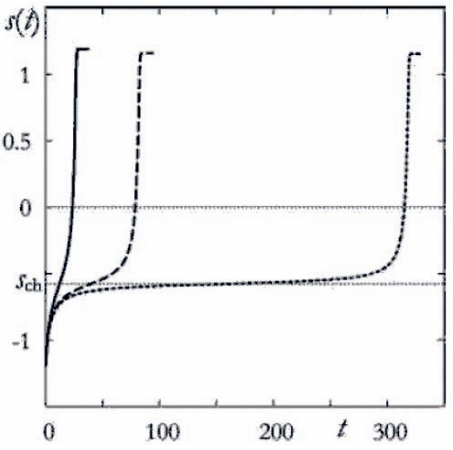

Fig. 2.5. Three orbits of the equation of motion (2.54) with $\epsilon = +$ for $H_0 > H_c$. The values of H_0 are set to be $H_0 = 0.386$ (dotted line), 0.4 (dashed line), and 0.5 (solid line), where all the initial conditions are chosen as $s_0 = s_-$ (adapted and modified from [OHF06]).

$$\dot{s} = s - s^3 + \epsilon H_0, \qquad (\epsilon = + \text{or} -) \tag{2.54}$$

for $H_0 > H_c$, i.e., $F(t)$ is fixed to be either $+H_0$ or $-H_0$. Equation (2.54) for $H_0 > H_c$ is integrated to yield

$$t = -\frac{1}{2(3s_\epsilon^2 - 1)} \ln \frac{(s - s_\epsilon)^2}{s^2 + s_\epsilon s + s_\epsilon^2 - 1} \frac{s_0^2 + s_\epsilon s_0 + s_\epsilon^2 - 1}{(s_0 - s_\epsilon)^2} \tag{2.55}$$
$$+ \frac{6 s_\epsilon}{2(3s_\epsilon^2 - 1)\sqrt{3s_\epsilon^2 - 4}} \left[\arctan\left(\frac{2s + s_\epsilon}{\sqrt{3s_\epsilon^2 - 4}} \right) - \arctan\left(\frac{2s_0 + s_\epsilon}{\sqrt{3s_\epsilon^2 - 4}} \right) \right],$$

where $s_0 = s(0)$ and s_ϵ has been defined in (2.48). Figure 2.5 displays three orbits given by (2.55) with $s_0 = s_-$ and $\epsilon = +$, which shows that $s(t)$ approaches s_+ in the limit $t \to \infty$. One observes that $s(t)$ stays for a long time in the vicinity of $s = -1/\sqrt{3}$ for H_0 slightly above H_c. The small region including the position $s = -1/\sqrt{3}$ is called the 'channel'. From the symmetry of the system, there also exists the channel near $s = 1/\sqrt{3}$ for $F(t) = -H_0$, as shown in Figure 2.3(b). Let us express the positions s_{ch} of the channels as

$$s_{ch} = \begin{cases} -1/\sqrt{3}, \text{ if } F(t) = +H_0 \\ +1/\sqrt{3}, \text{ if } F(t) = -H_0 \end{cases}.$$

The characteristic time τ_{ch} is then defined as the time span that the state point $s(t)$ passes through one of the channels for a constant $F(t)$, either $+H_0$ or $-H_0$. τ_{ch} can be estimated by integrating (2.54) around $s \simeq s_{ch}$ as follows. First, consider the case $F(t) = -H_0$. By setting $u(t) = s(t) - s_{ch}$ and assuming $|u| \ll s_{ch}$, (2.54) is approximated as [OHF06]

$$\dot{u} = -3s_{ch}u^2 - (H_0 - H_c).$$

This can be integrated to give

$$u(t) = -\sqrt{\frac{H_0 - H_c}{3s_{ch}}} \tan\left[\sqrt{3s_{ch}(H_0 - H_c)}t\right] \tag{2.56}$$

with the initial condition $u(0) = 0$. τ_{ch} is estimated by the condition $u(\tau_{ch}) = \infty$ and thus

$$\tau_{ch} = \frac{C}{(H_0 - H_c)^{1/2}}, \qquad C = \frac{\pi}{2\sqrt{3s_{ch}}}. \tag{2.57}$$

Let us next consider the process that the state point $s(t)$ passes through the channels under DMN. One finds that the time of passing through channels increases as H_0 approaches H_c. The MFPT $\bar{\tau}$ through channels was calculated in [HR83] by analyzing the master equation. In the present subsection, we will derive MFPT in terms of the time scales τ_f and τ_{ch} from a phenomenological viewpoint without use of the analysis made in [HR83].

The condition for passing through a channel is that $F(t)$ continues to take the identical value either $+H_0$ or $-H_0$ for time longer than τ_{ch}. For H_0 satisfying $\tau_f > \tau_{ch}$, we obtain $p(\tau_{ch}) = \mathrm{e}^{-\tau_{ch}/\tau_f} \simeq 1$, which implies that $F(t)$ almost always satisfies the condition for passing through the channel. Therefore, $\bar{\tau}$ in the case of $\tau_f > \tau_{ch}$ is nearly equivalent to τ_{ch}, i.e.,

$$\bar{\tau} \simeq \frac{C}{(H_0 - H_c)^{1/2}}. \tag{2.58}$$

In the case of $\tau_f \ll \tau_{ch}$, on the other hand, (2.46) gives $p(\tau_{ch}) \ll 1$. This fact implies that the probability that $F(t)$ continues to take the identical value for time longer than τ_{ch} is quite small and hence that $\bar{\tau}$ is much longer than τ_{ch} because it needs a long time to satisfy the condition for the state point to pass through the channel. $\bar{\tau}$ in the case of $\tau_f \ll \tau_{ch}$ is explicitly determined as follows. For a long $\bar{\tau}$, let us divide $\bar{\tau}$ into subintervals each of which has the time span τ_f. The divided individual time series are approximately independent of each other. Therefore, $\tau_f/\bar{\tau}$ is the probability that the state point passes through a channel once because $\bar{\tau}$ is MFPT through the channel. On the other hand, $p(\tau_{ch})$ is identical to the probability for $s(t)$ to pass through the channel once by definition of the probability. Therefore we get the relation $p(\tau_{ch}) \simeq \tau_f/\bar{\tau}$, which leads to

$$\bar{\tau}^{-1} \simeq \tau_f^{-1}\mathrm{e}^{-\tau_{ch}/\tau_f} = \tau_f^{-1} \exp\left[-\frac{C}{\tau_f(H_0 - H_c)^{1/2}}\right] \tag{2.59}$$

with the constant C defined in (2.57). This expression agrees with the result obtained in [HR83].

Equation (2.59) reveals that MFPT through the channel depends on $H_0 - H_c$ in a stretched exponential form for $\tau_f \ll \tau_{ch}$, and is quite different from the asymptotic form (2.58).

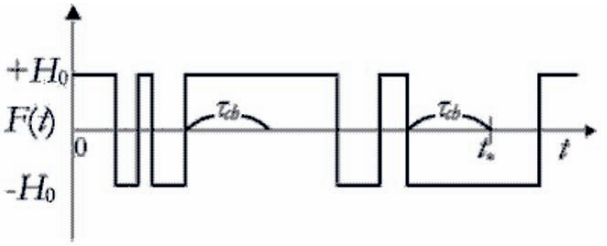

Fig. 2.6. Time series of $F(t)$ schematically indicating the time that $s(t)$ jumps from s_+ to s_-. $s(t)$ jumps at the time t_* in this case, where $F(t)$ takes the value $-H_0$ longer than τ_{ch} for the first time in the time series of $F(t)$ (adapted and modified from [OHF06]).

Phenomenological Analysis

In order to discuss statistical characteristics of the dynamics passing through the channels for $\tau_f \ll \tau_{ch}$, we here develop a phenomenological approach. The behaviors of $s(t)$ for which we attempt to model are first summarized. The initial condition of $s(t)$ is set to be in the vicinity of s_+. If a time interval of $F(t)$ satisfying the condition $F(t) = -H_0$ becomes longer than τ_{ch} for the first time, then $s(t)$ passes through s_{ch} and approaches s_- in the time interval. See Figure 2.6. The event in which $s(t)$ jumps from s_+ to s_- occurs only in this case. It should be noted that the jumps from $s(t) > 0$ ($s(t) < 0$) to $s(t) < 0$ ($s(t) > 0$) are approximately independent of subsequent jumps.

Let us discretize the time t in the form $t = k\Delta t$, ($k = 1, 2, 3, \cdots$) as a simple approach to develop the phenomenological analysis according to the process noted above, where Δt is a certain small time step. Then, τ_{ch} is discretized as $\tau_{ch} \equiv n_{ch}\Delta t$ with the corresponding integer n_{ch}. $F(t)$ is assumed to keep the same value for the interval Δt, which is denoted as $F_k = F(k\Delta t)$. The conditional probability p that F_{j+1} takes the same value as F_j is given by

$$p = \mathrm{e}^{-\Delta t/\tau_f}, \tag{2.60}$$

and the probability q that F_{k+1} is different from F_k is therefore given by

$$q = 1 - p. \tag{2.61}$$

The system is analyzed phenomenologically as follows:

- We introduce the variable s_k at a discretized time $k\Delta t$ which takes two values ± 1.
- s_k and F_k are initially set to $s_0 = +1$ and $F_0 = +H_0$, respectively.
- s_k jumps from $+1$ (-1) to -1 ($+1$) only if F_k continues to take the identical value $-H_0$ ($+H_0$) for a time interval longer than $n_{ch}\Delta t$.
- s_k does not jump from $+1$ (-1) to -1 ($+1$) even though F_k continues to take $+H_0$ ($-H_0$) for any time interval longer than $n_{ch}\Delta t$.

MFPT $\bar{\tau}$ through the Channel

We first derive the exact expression for the MFPT $\bar{\tau}$ through the channel with the phenomenological approach. In considering the time series having F_k, $\bar{\tau}$ is evaluated as

$$\bar{\tau} = \sum_l l\Delta t \sum_{0\le k\le l} g_{k,l}^{(n_{ch})} q^k p^{l-k} \Bigg|_{q=1-p} . \tag{2.62}$$

Here $g_{k,l}^{(n_{ch})}$ is the number of the time sequences $\{F_j\}$ for $0 \le j < l$ satisfying that F_j changed its value k times in each $\{F_j\}$ and s_j jumps from $+1$ to -1 for the first time at $t = l\Delta t$.

Equation (2.62) is, furthermore, rewritten as

$$\bar{\tau} = \hat{T} Q_{n_{ch}}(q,p)\Big|_{q=1-p}$$

with the differential operator $\hat{T}$ and the quantity $Q_{n_{ch}}(q,p)$ defined by

$$\hat{T} \equiv \Delta t \left(q\frac{\partial}{\partial q} + p\frac{\partial}{\partial p} \right) \tag{2.63}$$

and

$$Q_{n_{ch}}(q,p) \equiv \sum_l \sum_{0\le k\le l} g_{k,l}^{(n_{ch})} q^k p^{l-k}. \tag{2.64}$$

One should note that the q- and p-dependences in $Q_{n_{ch}}$ are crucial and that q and p are considered to be independent in (2.64).

The explicit form of $Q_{n_{ch}}(q,p)$ is then determined so as to satisfy the following conditions:

1. In considering any length of time series giving F_k, there exists a time interval of length n_{ch} in the last of the time series, where all the F_k take the same value $-H_0$, i.e., the condition that s_k jumps from $+1$ to -1 is satisfied.

2. The condition for s_k to jump from $+1$ to -1 is not satisfied before the last time interval.

One should note that the equality $Q_n(1-p,p) = 1$ holds for any n, because the time interval described above always exists somewhere in a long time series. Particularly, for $n = n_{ch}$, $Q_{n_{ch}}(1-p,p)$ is obviously equal to the probability that s_j changes its sign, which must be unity for $H_0 > H_c$.

The explicit form of $Q_n(q,p)$ is given by [OHF06]

$$Q_n(q,p) = \frac{(1-p)qp^{n-1}}{(1-p)^2 - q^2(1-p^{n-1})}, \tag{2.65}$$

where the condition $Q_n(1-p,p) = 1$ is easily confined. Applying the operator (2.63) to the explicit form (2.65) with $n = n_{ch}$ yields the relation

$$\bar{\tau} = \hat{T}Q_{n_{ch}}(q,p)\Big|_{q=1-p} = \Delta t \frac{2-p^{n_{ch}-1}}{(1-p)p^{n_{ch}-1}}$$
$$= \Delta t \frac{2-\mathrm{e}^{\Delta t/\tau_f}\mathrm{e}^{-\tau_{ch}/\tau_f}}{(1-\mathrm{e}^{-\Delta t/\tau_f})\mathrm{e}^{\Delta t/\tau_f}\mathrm{e}^{-\tau_{ch}/\tau_f}},$$

where the last equality is obtained by using Eqs. (2.60) and (2.61) with the relation $\tau_{ch} = n_{ch}\Delta t$. The exact expression of $\bar{\tau}$ is finally given by

$$\bar{\tau} = \tau_f \left(2\mathrm{e}^{\tau_{ch}/\tau_f} - 1\right) \tag{2.66}$$

in the limit of $\Delta t \to 0$ by keeping τ_{ch} constant. Equation (2.66) qualitatively agrees for $\tau_{ch}/\tau_f \gg 1$ with the result (2.59).

Distribution Function $P(\tau)$ for the Passage Time τ

The distribution function $P(\tau)$ for the passage time τ through the channel s_{ch} is determined by solving the equation

$$P(\tau) = \delta(\tau - \hat{T})Q_{n_{ch}}(q,p)\Big|_{q=1-p}, \tag{2.67}$$

where $\delta(x)$ is the delta function. The Laplace transform $\mathcal{L}[P](z)$ should be calculated in order to solve (2.67). By using the series expansion of $Q_{n_{ch}}(q,p)$ given by (2.64), the Laplace transform of $P(\tau)$ is obtained as

$$\mathcal{L}[P](z) \equiv \int_0^\infty \mathrm{e}^{-\tau z}P(\tau)d\tau = \mathrm{e}^{-z\hat{T}}Q_{n_{ch}}(q,p)\Big|_{q=1-p}$$
$$= \sum_l \sum_{0\le k\le l} g_{k,l}^{(n_{ch})}(\mathrm{e}^{-z\Delta t}q)^k(\mathrm{e}^{-z\Delta t}p)^{l-k}\Bigg|_{q=1-p}. \tag{2.68}$$

Equation (2.68) implies that $\mathcal{L}[P](z)$ can be obtained by replacing q and p by $\mathrm{e}^{-z\Delta t}q$ and $\mathrm{e}^{-z\Delta t}p$ in $Q_{n_{ch}}(q,p)$, respectively, i.e.,

$$\mathcal{L}[P(\tau)](z) = Q_{n_{ch}}(\mathrm{e}^{-z\Delta t}q, \mathrm{e}^{-z\Delta t}p)\big|_{q=1-p}. \tag{2.69}$$

Substituting the explicit form (2.65) for $n = n_{ch}$ into (2.69) yields the equation

$$\mathcal{L}[P(\tau)](z) = \frac{(\tau_f z+1)\mathrm{e}^{-(z+\tau_f^{-1})\tau_{ch}}}{\tau_f^2 z^2 + 2\tau_f z + \mathrm{e}^{-(z+\tau_f^{-1})\tau_{ch}}} \tag{2.70}$$

in the limit of $\Delta t \to 0$ by keeping τ_{ch} constant. By applying the inverse Laplace transform to (2.70), the distribution function $P(\tau)$ is analytically evaluated in the series expansion as

$$P(\tau) = \tau_f^{-1} \mathrm{e}^{-\tau/\tau_f} \sum_{k=0}^{\infty} \theta(t_{k+1}) \, \frac{(-x)^k}{k!} \frac{d^k}{dx^k} \cosh \sqrt{x} \Big|_{x=(t_{k+1})^2} =$$

$$\tau_f^{-1} \mathrm{e}^{-\tau/\tau_f} \Bigg[\theta(t_1) \cosh(t_1) - \theta(t_2) \frac{t_2 \sinh(t_2)}{2} + \theta(t_3) \frac{{t_3}^2 \cosh(t_3) - t_3 \sinh(t_3)}{8}$$

$$- \theta(t_4) \frac{{t_4}^3 \sinh(t_4) - 3{t_4}^2 \cosh(t_4) + 3t_4 \sinh(t_4)}{48} + \cdots \Bigg], \tag{2.71}$$

where $t_k(\tau) \equiv (\tau - k\tau_{ch})/\tau_f$ and $\theta(t)$ is the Heaviside function defined by

$$\theta(t) = \begin{cases} 1 \text{ for } t \geq 0 \\ 0 \text{ for } t < 0 \end{cases} . \tag{2.72}$$

For details of the derivation of (2.71), see [OHF06].

Let us suppose to truncate the expansion (2.71) at $k = k_c$ for an arbitrary k_c. It should be noted that (2.71) gives the exact distribution for $0 < \tau < k_c \tau_{ch}$ even though the truncation is executed, since all the terms individually include $\theta(t_k)$ and so the terms for $k > k_c$ do not contribute to $P(\tau)$ for $\tau < k_c \tau_{ch}$. The analytical result (2.71) is compared with the numerically evaluated distribution. One observes that the phenomenological analysis quantitatively explains the statistical property of passing through the channels. The characteristics obtained from the figure are summarized as follows [OHF06]:

1. There exists a region where $P(\tau) = 0$ for $\tau < \tau_{ch}$, which presents the minimal time of passing through the channels.

2. $P(\tau)$ decreases exponentially for $\tau \gg \tau_{ch}$, $P(\tau) \propto \mathrm{e}^{-\alpha\tau}$ with a constant α.

3. The rate α increases as τ_f is increased. This tendency is consistent with the fact that the probability of passing through channels increases as τ_f is increased since DMN will often continue to take an identical value longer than τ_{ch}.

On the other hand, the expansion (2.71) disagrees with the correct value in an exponential way for $\tau > k_c \tau_{ch}$. Let us try to obtain the asymptotic solution of $P(\tau)$ for $\tau \gg \tau_{ch}$. We have [OHF06]

$$\langle \mathrm{e}^{-z(\tau - \tau_{ch})} \rangle \simeq \frac{1}{1 + (\bar{\tau} - \tau_{ch}) z} \quad \text{for } |z| \ll \tau_{ch}^{-1}, \tag{2.73}$$

where $\bar{\tau}$ is MFPT given in (2.66). The inverse Laplace transform of (2.73) is straightforwardly calculated to give

$$P(\tau) \simeq \frac{1}{\bar{\tau} - \tau_{ch}} \exp\left(-\frac{\tau - \tau_{ch}}{\bar{\tau} - \tau_{ch}} \right),$$

which reveals that $P(\tau)$ decreases exponentially with the damping rate $\alpha = (\bar{\tau} - \tau_{ch})^{-1}$ for $\tau \gg \tau_{ch}$.

Phenomenological Fourier Spectrum

We derive the Fourier spectrum of a time series $s(t)$ by the phenomenological analysis to focus on the dynamical characteristics in the SRM phase. The Fourier spectrum $I_x(\omega)$ is defined by

$$I_x(\omega) = \lim_{T\to\infty} \frac{1}{T} \left\langle \left| \int_0^T x(t) \mathrm{e}^{-i\omega t} dt \right|^2 \right\rangle, \tag{2.74}$$

i.e., the ensemble average of the Fourier transform of a time series $x(t)$.

Let us first consider $s_0(t) \equiv \mathrm{sgn}[s(t)]$. Then the time series $s_0(t)$ is expressed as

$$s_0(t) = (-1)^{n-1}, \qquad \text{for} \quad t_{n-1} \le t < t_n \tag{2.75}$$

with $n \ge 1$, where t_n denotes the nth time to cross zero for $s(t)$. Hereafter, t_0 is set to be zero without loss of generality. By identifying that $\tau_n \equiv t_n - t_{n-1}$ is independently distributed according to (2.71), one obtains the Fourier spectrum of $s_0(t)$ by the phenomenological analysis shown (see [OHF06]), in the form

$$I_{s_0}(\omega) = \frac{4}{\bar{\tau}\omega^2} \Re \left(\frac{1 - \langle \mathrm{e}^{-i\omega\tau_n} \rangle}{1 + \langle \mathrm{e}^{-i\omega\tau_n} \rangle} \right) = \frac{4}{\bar{\tau}\omega^2} \frac{1 - |\langle \mathrm{e}^{-i\omega\tau_n} \rangle|^2}{|1 + \langle \mathrm{e}^{-i\omega\tau_n} \rangle|^2} \tag{2.76}$$

where $\Re(X)$ represents the real part of X, and $\lim_{N\to\infty} \frac{t_N}{N} = \langle \tau_n \rangle = \bar{\tau}$ is used.

Substituting the explicit form of $\langle \mathrm{e}^{-i\omega\tau_n} \rangle$ given in (2.70) with $z = i\omega$ into (2.76) yields

$$I_{s_0}(\omega) = \left(\frac{4\tau_f}{\bar{\tau}\omega} \right) \frac{\omega^3\tau_f^3 + (4 - \mathrm{e}^{-2\tau_{ch}/\tau_f})\omega\tau_f - 2\mathrm{e}^{-\tau_{ch}/\tau_f}(\omega\tau_f \cos\omega\tau_{ch} + 2\sin\omega\tau_{ch})}{(4 + \omega^2\tau_f^2)(\omega^2\tau_f^2 - 2\omega\tau_f \mathrm{e}^{-\tau_{ch}/\tau_f} \sin\omega\tau_{ch} + \mathrm{e}^{-2\tau_{ch}/\tau_f})}. \tag{2.77}$$

The above result is confirmed by comparing with the numerically evaluated Fourier spectrum for the normalized time series $s_0(t)$ (see [OHF06]).

Let us finally modify the phenomenological analysis which is compatible with the numerically evaluated spectrum of the original time series $s(t)$ without normalization. Instead of (2.75), let us define

$$\tilde{s}(t) = (-1)^{n-1}[1 - a(t - t_{n-1})] \quad \text{for } t_{n-1} \le t < t_n$$

with $n \ge 1$, where $a(\Delta t)$ incorporates the wave form of the time series passing through the channel and is assumed to be $a(\Delta t) = 0$ for $\Delta t > \tau_{ch}$. Note that by setting $a(\Delta t) = 0$ also for $\Delta t \le \tau_{ch}$ the result of original phenomenological analysis is recovered. As shown in [OHF06], the Fourier spectrum $I_{\tilde{s}}(\omega)$ for $\tilde{s}(t)$ as a modification to $I_{s_0}(\omega)$ is obtained in the form

$$I_{\tilde{s}}(\omega) = I_{s_0}(\omega) \frac{1 + |\hat{a}(\omega)|^2 + 2\Re[\hat{a}(\omega)]}{4}, \tag{2.78}$$

where

$$\hat{a}(\omega) \equiv 1 - i\omega \int_0^{\tau_{ch}} a(t)\mathrm{e}^{-i\omega t} dt.$$

By approximating as $a(\Delta t) = 1 + |s_{ch}|$ for $0 < \Delta t < \tau_{ch}$, (2.78) reduces to

$$I_{\tilde{s}}(\omega) = I_{s_0}(\omega) \left(\frac{1 + s_{ch}^2}{2} + \frac{1 - s_{ch}^2}{2} \cos \omega \tau_{ch} \right). \tag{2.79}$$

Fore more technical details, see [OHF06].

2.1.11 Phase Transition in a Reaction–Diffusion System

Recently 1D *reaction–diffusion systems* have received much attention because they show a variety of interesting critical phenomena such as *non–equilibrium phase transitions* [Sch01]. A simple system of this type, which has been studied widely in related literatures, is the Asymmetric Simple Exclusion Process (ASEP) [DEH93]. In this 2–states model the particles are injected from the left site of an open discreet lattice of length L. They diffuse in the system and at the end of the lattice are extracted from the system. It is known that depending on the injection and the extraction rates the ASEP shows different boundary induced phase transitions. Non–equilibrium phase transition may also happen in the systems with *non–conserving dynamics*. For instance in [EKL02] the authors investigate a 3–states model consists of two species of particles besides vacancies on a lattice with ring geometry. The dynamics of this model consists of diffusion, creation and annihilation of both species of the particles. They have found that the phase diagram of the model highly depends on the annihilation rate of the particles. By changing the annihilation rate of the particles, the system transfers from a maximal current phase to a fluid phase. The density of the vacancies changes discontinuously from one phase to the other phase.

In this section, following [JG07], we study a reaction–diffusion model on a discrete lattice of length L with periodic boundary condition. Besides the vacancies there are two different types of particles in the system. Throughout this paper the vacancies and the particles are denoted by E, A and B. The dynamics of the system is not conserving. The particles of type A and type B hop to the left and to the right respectively. The total number of particles of type B is a conserved quantity and assumed to be equal to M. The density of these particles is defined as $\rho_B = \frac{M}{L}$. In contrast, the total number of particles of type A is not a conserved quantity due to the creation and annihilation of them. Only the nearest neighbors interactions are allowed and the model evolves through the following processes

$$\begin{aligned}
&A\emptyset \longrightarrow \emptyset A \quad \text{with rate } 1,\\
&\emptyset B \longrightarrow B\emptyset \quad \text{with rate } 1,\\
&AB \longrightarrow BA \quad \text{with rate } 1,\\
&A\emptyset \longrightarrow \emptyset\emptyset \quad \text{with rate } \omega,\\
&\emptyset\emptyset \longrightarrow A\emptyset \quad \text{with rate } 1.
\end{aligned} \tag{2.80}$$

As can be seen the parameter ω determines the annihilation rate for the particles of type A which besides the number of the particles of type B i.e., ρ_B are the free parameters of the model. One should note that the annihilation in our model only takes place for one species of particles. Our main aim in the present work is to study the phase diagram of the model in terms of ω and the density of the B particles.

In our model if one starts with a lattice without any vacancies the dynamics of the model prevents it from evolving into other configurations consisting of vacancies. In this case the system remains in its initial configuration and the steady state of the system is trivial. In order to study the non-trivial case we consider those configurations which have at least one vacancy. In order to find the stationary probability distribution function of the system we apply the *Matrix Product Formalism* (MPF) first introduced in [DEH93] and then generalized in [KS97]. According to this formalism the stationary probability for any configuration of a system with periodic boundary condition is proportional to the trace of product of non–commuting operators which satisfy a quadratic algebra. In our model we have three different states at each site of the lattice associated with the presence of vacancies, the A particles and the B particles. We assign three different operators $\mathbf{E}$, $\mathbf{A}$ and $\mathbf{B}$ to each state. Now the unnormalized steady state probability of a configuration $\mathcal{C}$ is given by

$$P(\mathcal{C}) = \frac{1}{\mathcal{Z}_L} Tr[\prod_{i=1}^{L} \mathbf{X}_i], \tag{2.81}$$

in which $\mathbf{X}_i = \mathbf{E}$ if the site i is empty, $\mathbf{X}_i = \mathbf{A}$ if the site i is occupied by a particle of type A and $\mathbf{X}_i = \mathbf{B}$ if it is occupied by a particle of type B. The normalization factor $\mathcal{Z}_L$ in the denominator of (2.81) is called the partition function of the system and is given by the sum of unnormalized weights of all accessible configurations. By applying the MPF one finds the following quadratic algebra for our model [JG07]

$$\begin{aligned} &\mathbf{AB} = \mathbf{A} + \mathbf{B}, \\ &\mathbf{AE} = \mathbf{E}, \\ &\mathbf{EB} = \mathbf{E}, \\ &\mathbf{E}^2 = \omega \mathbf{E}. \end{aligned} \tag{2.82}$$

Now by defining

$$\mathbf{E} = \omega |V\rangle\langle W| \qquad \text{in which} \qquad \langle W|V\rangle = 1,$$

one can simply find

$$\begin{aligned} &\mathbf{AB} = \mathbf{A} + \mathbf{B}, \\ &\mathbf{A}|V\rangle = |V\rangle, \\ &\langle W|\mathbf{B} = \langle W|, \\ &\mathbf{E} = \omega |V\rangle\langle W|. \end{aligned} \tag{2.83}$$

The first three relations in (2.83) make a quadratic algebra which is well known in the related literatures. It is the quadratic algebra of the ASEP when the

boundary rates are equal to one.[4] This algebra has an infinite dimensional representation given by the following matrices and vectors

$$\mathbf{A} = \begin{pmatrix} 1\,1\,0\,0\cdots \\ 0\,1\,1\,0 \\ 0\,0\,1\,1 \\ 0\,0\,0\,1 \\ \vdots \qquad \ddots \\ \end{pmatrix}, \qquad \mathbf{B} = \mathbf{A}^T,$$

$$|V\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ \vdots \end{pmatrix}, \qquad \langle W| = |V\rangle^T,$$

in which T stands for transpose. In order to find the phase structure of the system one can calculate the generating function of the partition function of the system and study its singularities. In what follows we first calculate the grandcanonical partition function of the system $\mathcal{Z}_L(\xi)$ by introducing a fugacity ξ for particles of type B. We then fix the fugacity of the B particles using the following relation

$$\rho_B = \lim_{L\to\infty} \frac{\xi}{L} \frac{\partial \ln \mathcal{Z}_L(\xi)}{\partial \xi}. \tag{2.84}$$

The grandcanonical *partition function* of the system $\mathcal{Z}(\xi)$ can now be calculated from (2.81) and is given by [JG07]

$$\mathcal{Z}_L(\xi) = \sum_{\mathcal{C}} Tr[\prod_{i=1}^{L} \mathbf{X}_i] = Tr[(\mathbf{A} + \xi\mathbf{B} + \mathbf{E})^L - (\mathbf{A} + \xi\mathbf{B})^L]. \tag{2.85}$$

One should note that the operator $\mathbf{B}$ in (2.81) is replaced with the operator $\xi\mathbf{B}$ in which ξ should be fixed using (2.84). As we mentioned above the stationary state of the system without vacancies is a trivial one, therefore in (2.85) we have considered those configurations with at least one vacancy. The generating function for $\mathcal{Z}_L(\xi)$ can now be calculated using (2.83). Using the same procedure introduced in [RSS00] and after some straightforward algebra one finds

$$\mathcal{G}(\xi,\lambda) = \sum_{L=0}^{\infty} \lambda^{L-1} \mathcal{Z}_L(\xi) = \frac{\frac{d}{d\lambda}\omega U(\xi,\lambda)}{1 - \omega U(\xi,\lambda)}, \tag{2.86}$$

in which

[4] The operator $\mathbf{A}$ and $\mathbf{B}$ should be regarded as the operators associated with the particles and vacancies respectively.

$$U(\xi,\lambda)=\sum_{L=0}^{\infty}\lambda^{L+1}\langle W|(\mathbf{A}+\xi\mathbf{B})^{L}|V\rangle.$$

The convergence radius of the formal series (2.86) which is the absolute value of its nearest singularity to the origin can be written as

$$R(\xi)=\lim_{L\to\infty}\mathcal{Z}_{L}(\xi)^{\frac{-1}{L}}.$$

This is also the inverse of the largest eigenvalue of $\mathcal{Z}_L(\xi)$. In the large L limit using (2.84) this results in the following relation

$$\rho_B=\xi\frac{\partial}{\partial\xi}\ln\frac{1}{R(\xi)}. \tag{2.87}$$

Therefore one should only find the singularities of (2.86) and decide in which region of the phase diagram which singularity is the smallest one. In order to find the singularities of (2.86) one should first calculate $U(\xi,\lambda)$. This can easily be done by noting that the matrix $\mathbf{A}+\xi\mathbf{B}$ defined as

$$\mathbf{A}+\xi\mathbf{B}=\begin{pmatrix}1+\xi & 1 & 0 & 0 & \cdots\\ \xi & 1+\xi & 1 & 0 & \\ 0 & \xi & 1+\xi & 1 & \\ 0 & 0 & \xi & 1+\xi & \\ \vdots & & & & \ddots\end{pmatrix},$$

satisfy the following *eigenvalue relation*

$$(\mathbf{A}+\xi\mathbf{B})|\theta\rangle=(1+\xi+2\sqrt{\xi}Cos(\theta))|\theta\rangle, \tag{2.88}$$

for $-\pi\leq\theta\leq\pi$ in which we have defined [JG07]

$$|\theta\rangle=\begin{pmatrix}Sin(\theta)\\ \xi^{1/2}Sin(2\theta)\\ \xi Sin(3\theta)\\ \xi^{3/2}Sin(4\theta)\\ \xi^{2}Sin(5\theta)\\ \vdots\end{pmatrix}.$$

Considering the fact that

$$\int_{-\pi}^{\pi}\frac{d\theta}{\pi}Sin(\theta)Sin(n\theta)=\delta_{1,n},$$

one can easily see that

$$|V\rangle = \int_{-\pi}^{\pi} \frac{d\theta}{\pi} Sin(\theta)|\theta\rangle. \tag{2.89}$$

Now using (2.88)–(2.89) and the fact that $\langle W|\theta\rangle = Sin(\theta)$ one can easily show that $U(\xi,\lambda)$ is given by

$$U(\xi,\lambda) = \int_{-\pi}^{\pi} \frac{d\theta}{\pi} \frac{\lambda Sin(\theta)^2}{1-\lambda(1+\xi+2\sqrt{\xi}Cos(\theta))} \tag{2.90}$$

which is valid for $\lambda < \frac{1}{(1+\sqrt{\xi})^2}$. The integral (2.90) can easily be calculated using the Cauchy residue theorem by noting that it has three poles in the complex plane. Two of them are inside the contour which is a circle of unite radius around the origin and one is outside it. After some calculations one finds

$$U(\xi,\lambda) = \frac{1-\lambda-\lambda\xi-\sqrt{(\lambda+\lambda\xi-1)^2-4\lambda^2\xi}}{2\lambda\xi}.$$

Now that $U(\xi,\lambda)$ is calculated one can simply find the singularities of (2.86). It turns out that (2.86) has two different kinds of singularities: a simple root singularity $R_1 = \frac{\omega}{(1+\omega)(\omega+\xi)}$ which come from the denominator of (2.86) and a square root singularity $R_2 = \frac{1}{(1+\sqrt{\xi})^2}$. Therefore the model has two different phases. The relation between the density of B particles and their fugacity in each phase should be obtained from (2.87). In terms of the density of the B particles ρ_B we find

$$R_1 = \frac{1-\rho_B}{1+\omega} \qquad \text{and} \qquad R_2 = (1-\rho_B)^2.$$

Two different scenarios might happen: defining $\omega_c = \frac{\rho_B}{1-\rho_B}$ we find that for $\omega > \omega_c$ the nearest singularity to the origin is R_1 and for $\omega < \omega_c$ it is R_2. The density of the vacancies can be calculated quit similar to that of the B particles. It is given by

$$\rho_E = \omega \frac{\partial}{\partial\omega} \ln \frac{1}{R(\xi)}, \tag{2.91}$$

in which $R(\xi)$ is again the nearest singularity to the origin in each phase. The density of the A particles is in turn $\rho_A = 1-\rho_B-\rho_E$. Let us now investigate the current of the particles in each phase. Noting that the configurations without vacancies are inaccessible, the particle current for each species is obtained to be [JG07]

$$J_{\mathbf{A}} = \frac{Tr[(\xi\mathbf{AB}+\mathbf{AE})(\mathbf{A}+\xi\mathbf{B}+\mathbf{E})^{L-2} - (\xi\mathbf{AB})(\mathbf{A}+\xi\mathbf{B})^{L-2}]}{Tr[(\mathbf{A}+\xi\mathbf{B}+\mathbf{E})^L - (\mathbf{A}+\xi\mathbf{B})^L]},$$
$$J_{\mathbf{B}} = \frac{Tr[(\xi\mathbf{AB}+\xi\mathbf{EB})(\mathbf{A}+\xi\mathbf{B}+\mathbf{E})^{L-2} - (\xi\mathbf{AB})(\mathbf{A}+\xi\mathbf{B})^{L-2}]}{Tr[(\mathbf{A}+\xi\mathbf{B}+\mathbf{E})^L - (\mathbf{A}+\xi\mathbf{B})^L]}.$$

These relations can be simplified in the thermodynamic limit and one finds

$$J_A = R(\xi)(1 + (\xi - 1)\rho_A),$$
$$J_B = R(\xi)(\xi + (1 - \xi)\rho_B).$$

As we mentioned above for $\omega < \omega_c$ the nearest singularity to the origin is always R_2. The particle currents in this case are equal and we find

$$J_A = J_B = \rho_B(1 - \rho_B).$$

In contrast for $\omega > \omega_c$ the nearest singularity to the origin is always R_1 and it turns out that the currents are not equal. We find $J_A = \frac{\omega}{(1+\omega)^2}$ and $J_B = \rho_B(1 - \rho_B)$. In what follows we bring the summery of the results concerning the phase structure of the system

$$\text{for } \omega < \omega_c \quad \begin{cases} \rho_A = 1 - \rho_B & J_A = \rho_B(1 - \rho_B) \\ \rho_B = \rho_B & J_B = \rho_B(1 - \rho_B) \\ \rho_E = 0 \end{cases} \qquad \text{and}$$

$$\text{for } \omega > \omega_c \quad \begin{cases} \rho_A = \frac{1}{1+\omega} & J_A = \frac{\omega}{(1+\omega)^2} \\ \rho_B = \rho_B & J_B = \rho_B(1 - \rho_B) \\ \rho_E = \frac{\omega}{1+\omega} - \rho_B \end{cases} .$$

As can be seen for $\omega < \omega_c$ the density of vacancies is equal to zero which means there are only A and B particles on the lattice. Since the density of the B particles is fixed and equal to ρ_B the density of A particles should be $1 - \rho_B$. In this case, according to (2.80), both A and B particles have simply ASEP dynamics and therefore their currents should be of the form $\rho(1 - \rho)$ and that $J_A = J_B$. This is in quite agreement with our calculations for J_A and J_B. On the other hand for $\omega > \omega_c$ the density of vacancies on the lattice is no longer zero. At the transition point ω_c the density of the vacancies is zero but it increases linearly in this phase. In terms of the density of the vacancies the phase transition is a continuous transition. In this phase for $\omega < 1$ we always have $J_A > J_B$ while for $\omega > 1$ we have

$$J_A > J_B \qquad \text{for} \qquad \rho_B < \frac{1}{1+\omega}$$
$$\text{and} \qquad J_A < J_B \qquad \text{for} \qquad \frac{1}{1+\omega} < \rho_B < \frac{\omega}{1+\omega}.$$

In this paper we studied a 3–states model consists of A and B particles besides the vacancies. The A particles are created and annihilated which is controlled by ω while the B particles only diffuse on the lattice and have a fixed density ρ_B. We found that the system have two phases depending on ρ_B and ω. The current of the B particles is always a constant $\rho_B(1 - \rho_B)$ throughout the phase diagram while for the A particles it is given by different expressions in each phase. As a generalization one could also consider a more general process

$$
\begin{aligned}
A\emptyset &\longrightarrow \emptyset A \quad \text{with rate } \alpha,\\
\emptyset B &\longrightarrow B\emptyset \quad \text{with rate } \beta,\\
AB &\longrightarrow BA \quad \text{with rate } 1,\\
A\emptyset &\longrightarrow \emptyset\emptyset \quad \text{with rate } \lambda,\\
\emptyset\emptyset &\longrightarrow A\emptyset \quad \text{with rate } \lambda',
\end{aligned}
$$

with the quadratic algebra given by [JG07]

$$
\begin{aligned}
&\mathbf{AB} = \mathbf{A} + \mathbf{B},\\
&\alpha\ \mathbf{A}|V\rangle = |V\rangle,\\
&\beta\ \langle W|\mathbf{B} = \langle W|,\\
&\mathbf{E} = \tfrac{\lambda}{\lambda'\alpha}|V\rangle\langle W|.
\end{aligned}
$$

and apply the same approach used in present section to study its phase diagram.

2.1.12 Phase Transition in Negotiation Dynamics

Statistical physics has recently proved to be a powerful framework to address issues related to the characterization of the collective social behavior of individuals, such as culture dissemination, the spreading of linguistic conventions, and the dynamics of opinion formation [Wei00].

According to the 'herding behavior' described in sociology [Cha03], processes of opinion formation are usually modelled as simple collective dynamics in which the agents update their opinions following local majority [Gla63] or imitation rules [Lig85]. Starting from random initial conditions, the system self–organizes through an ordering process eventually leading to the emergence of a global consensus, in which all agents share the same opinion. Deviations from purely herding behavior are considered by introducing a certain level of noise. In analogy with kinetic Ising models and contact processes [Odo04], the presence of noise can induce non–equilibrium phase transitions from the *consensus* state to disordered configurations, in which more than one opinion is present.

The principle of 'bounded confidence' [DNA00], on the other hand, consists in enabling interactions only between agents that share already some cultural features (defined as discrete objects) [Axe97] or with not too different opinions (in a continuous space) [DNA00, BKR03]. The overall behavior of the system depends on the method used to discriminate 'different' and 'similar' opinions. By tuning some threshold parameter, transitions are observed concerning the number of opinions surviving in the (frozen) final state. This can be a situation of *consensus*, in which all agents share the same opinion, *polarization*, in which a finite number of groups with different opinions survive, or *fragmentation*, with a final number of opinions scaling with the system size. For instance, in the *Axelrod model* [Axe97] a consensus–to–fragmentation transition occurs as the variability of cultural features is increased [CMV00].

In this section, following [BDB07], we present a model of *opinion dynamics* in which a consensus–polarization–fragmentation non–equilibrium phase transition is driven by external noise, intended as an 'irresolute attitude' of the agents in making decisions. The primary attribute of the model is that it is based on a negotiation process, in which memory and feedback play a central role. Moreover, apart from the consensus state, no configuration is frozen: the stationary states with several coexisting opinions are still dynamical, in the sense that the agents are still able to evolve, in contrast to the Axelrod model [Axe97].

Let us consider a population of N agents, each one endowed with a memory, in which an a priori undefined number of opinions can be stored. In the initial state, agents memories are empty. At each time step, an ordered pair of neighboring agents is randomly selected. This choice is consistent with the idea of *directed attachment* in the socio–psychological literature (see for instance [Fri90]). The negotiation process is described by a local pairwise interaction rule: a) the first agent selects randomly one of its opinions (or creates a new opinion if its memory is empty) and conveys it to the second agent; b) if the memory of the latter contains such an opinion, with probability β the two agents update their memories erasing all opinions except the one involved in the *interaction* (*agreement*), while with probability $1 - \beta$ nothing happens; c) if the memory of the second agent does not contain the uttered opinion, it adds such an opinion to those already stored in its *memory* (*learning*). Note that, in the special case $\beta = 1$, the negotiation rule reduces to the Naming Game rule [BFC06], a model used to describe the emergence of a communication system or a set of linguistic conventions in a population of individuals. In our modelling the parameter β plays roughly the same role as the *probability of acknowledged influence* in the socio–psychological literature [Fri90]. Furthermore, as already stated for other models [Cas05], when the system is embedded in heterogeneous topologies, different pair selection criteria influence the dynamics. In the *direct strategy*, the first agent is picked up randomly in the population, and the second agent is randomly selected among its neighbors. The opposite choice is called *reverse strategy*; while the *neutral strategy* consists in randomly choosing a link, assigning it an order with equal probability.

At the beginning of the dynamics, a large number of opinions is created, the total number of different opinions growing rapidly up to $\mathcal{O}(N)$. Then, if β is sufficiently large, the number of opinions decreases until only one is left and the consensus state is reached (as for the Naming Game in the case $\beta = 1$). In the opposite limit, when $\beta = 0$, opinions are never eliminated, therefore the only possible stationary state is the trivial state in which every agent possesses all opinions. Thus, a non–equilibrium phase transition is expected for some critical value β_c of the parameter β governing the update efficiency. In order to find β_c, we exploit the following general stability argument. Let us consider the consensus state, in which all agents possess the same unique opinion, say A. Its stability may be tested by considering a situation in which

A and another opinion, say B, are present in the system: each agent can have either only opinion A or B, or both (AB state). The critical value β_c is provided by the threshold value at which the perturbed configuration with these three possible states does not converge back to consensus.

The simplest assumption in modelling a population of agents is the homogeneous mixing (i.e., mean–field (MF) approximation), where the behavior of the system is completely described by the following evolution equations for the densities n_i of agents with the opinion i [BDB07]

$$\begin{aligned} dn_A/dt &= -n_A n_B + \beta n_{AB}^2 + \frac{3\beta - 1}{2} n_A n_{AB}, \\ dn_B/dt &= -n_A n_B + \beta n_{AB}^2 + \frac{3\beta - 1}{2} n_B n_{AB}, \end{aligned} \tag{2.92}$$

and $n_{AB} = 1 - n_A - n_B$. Imposing the steady state condition $\dot{n}_A = \dot{n}_B = 0$, we get three possible solutions: 1) $n_A = 1$, $n_B = 0$, $n_{AB} = 0$; 2) $n_A = 0$, $n_B = 1$, $n_{AB} = 0$; and 3) $n_A = n_B = b(\beta)$, $n_{AB} = 1 - 2b(\beta)$ with $b(\beta) = \frac{1+5\beta-\sqrt{1+10\beta+17\beta^2}}{4\beta}$ (and $b(0) = 0$). The study of the solutions' stability predicts a phase transition at $\beta_c = 1/3$. The maximum non–zero eigenvalue of the linearized system around the consensus solution becomes indeed positive for $\beta < 1/3$, i.e., the consensus becomes unstable, and the population polarizes in the $n_A = n_B$ state, with a finite density of undecided agents n_{AB}. The model therefore displays a first order non–equilibrium transition between the *frozen* absorbing consensus state and an *active polarized state*, in which global observables are stationary on average, but not frozen, i.e., the population is split in three dynamically evolving parts (with opinions A, B, and AB), whose densities fluctuate around the average values $b(\beta)$ and $1 - 2b(\beta)$.

We have checked the predictions of (2.92) by numerical simulations of N agents interacting on a complete graph [BDB07]. The convergence time t_{conv} required by the system to reach the consensus state indeed diverges at $\beta_c = 1/3$, with a power–law behavior $(\beta - \beta_c)^{-a}$, $a \simeq 0.3$. [5] Very interestingly however, the analytical and numerical analysis of (2.92) predicts that the relaxation time diverges instead as $(\beta - \beta_c)^{-1}$. This apparent discrepancy arises in fact because (2.92) consider that the agents have at most two different opinions at the same time, while this number is unlimited in the original model (and in fact diverges with N). Numerical simulations reproducing the two opinions case allow to recover the behavior of t_{conv} predicted from (2.92). We have also investigated the case of a finite number m of opinions available to the agents. The analytical result $a = 1$ holds also for $m = 3$ (but analytical analysis for larger m becomes out of reach), whereas preliminary numerical simulations performed for $m = 3, 10$ with the largest reachable population size

[5] The low value of a, which moreover slightly decreases as the system size increases, does not allow to exclude a logarithmic divergence. This issue deserves more investigations that we leave for future work.

($N = 10^6$) lead to an exponent $a \simeq 0.74 \div 0.8$. More extensive and systematic simulations are in order to determine the possible existence of a series of universality classes varying the memory size for the agents. In any case, the models with finite (m opinions) or unlimited memory define at least two clearly different universality classes for this non–equilibrium phase transition between consensus and polarized states (see [CG94] for similar findings in the framework of non–equilibrium q-state systems).

Also, the transition at β_c is only the first of a series of transitions: when decreasing $\beta < \beta_c$, a system starting from empty initial conditions self-organizes into a fragmented state with an increasing number of opinions. In principle, this can be shown analytically considering the mean-field evolution equations for the partial densities when $m > 2$ opinions are present, and studying, as a function of β, the sign of the eigenvalues of a $(2^m - 1) \times (2^m - 1)$ stability matrix for the stationary state with m opinions. For increasing values of m, such a calculation becomes rapidly very demanding, thus we limit our analysis to the numerical insights, from which we also get that the number of residual opinions in the fragmented state follows the *exponential law* [BDB07]

$$m(\beta) \propto \exp\left[(\beta_c - \beta)/C\right],$$

where C is a constant depending on the initial conditions (not shown).

We now extend the present analysis to more general interactions topologies, in which agents are placed on the vertices of a network, and the edges define the possible interaction patterns [BDB07]. When the network is a homogeneous random one (Erdös–Rényi (ER) graph [ER59]), the degree distribution is peaked around a typical value $\langle k \rangle$, and the evolution equations for the densities when only two opinions are present provide the same transition value $\beta_c = 1/3$ and the same exponent -1 for the divergence of t_{conv} as in MF.

Since any real negotiation process takes place on social groups, whose topology is generally far from being homogeneous, we have simulated the model on various uncorrelated heterogeneous networks (using the *uncorrelated configuration model* (UCM) [CBP05]), with power–law degree distributions $P(k) \sim k^{-\gamma}$ with exponents $\gamma = 2.5$ and $\gamma = 3$.

Very interestingly, the model still presents a consensus-polarization transition, in contrast with other *opinion–dynamics* models, like for instance the Axelrod model [KET03], for which the transition disappears for heterogeneous networks in the thermodynamic limit.

To understand these numerical results, we analyze, as for the fully connected case, the evolution equations for the case of two possible opinions. Such equations can be written for general correlated complex networks whose topology is completely defined by the degree distribution $P(k)$, i.e., the probability that a node has degree k, and by the degree–degree conditional probability $P(k'|k)$ that a node of degree k' is connected to a node of degree k (*Markovian networks*). Using partial densities

$$n_A^k = N_A^k/N_k, n_B^k = N_B^k/N_k \qquad \text{and} \qquad n_{AB}^k = N_{AB}^k/N_k,$$

i.e., the densities on classes of degree k, one derives mean–field type equations in analogy with epidemic models. Let us consider for definiteness the neutral pair selection strategy, the equation for n_A^k is in this case [BDB07]

$$\begin{aligned}\frac{dn_A^k}{dt} &= -\frac{1}{\langle k\rangle} n_A^k k \sum_{k'} P(k'|k) n_B^{k'} - \frac{1}{2\langle k\rangle} n_A^k k \sum_{k'} P(k'|k) n_{AB}^{k'} \\ &+ \frac{3\beta}{2\langle k\rangle} k n_{AB}^k \sum_{k'} P(k'|k) n_A^{k'} + 2\frac{\beta}{2\langle k\rangle} k n_{AB}^k \sum_{k'} P(k'|k) n_{AB}^{k'}, \\ \frac{dn_A^k}{dt} &= -\frac{k n_A^k}{\langle k\rangle} \sum_{k'} P(k'|k) n_B^{k'} - \frac{k n_A^k}{2\langle k\rangle} \sum_{k'} P(k'|k) n_{AB}^{k'} \\ &+ \frac{3\beta k n_{AB}^k}{2\langle k\rangle} \sum_{k'} P(k'|k) n_A^{k'} + \frac{\beta k n_{AB}^k}{\langle k\rangle} \sum_{k'} P(k'|k) n_{AB}^{k'} \ ,\end{aligned} \tag{2.93}$$

and similar equations hold for n_B^k and n_{AB}^k. The first term corresponds to the situation in which an agent of degree k' and opinion B chooses as second actor an agent of degree k with opinion A. The second term corresponds to the case in which an agent of degree k' with opinions A and B chooses the opinion B, interacting with an agent of degree k and opinion A. The third term is the sum of two contributions coming from the complementary interaction; while the last term accounts for the increase of agents of degree k and opinion A due to the interaction of pairs of agents with AB opinion in which the first agent chooses the opinion A.

Now, let us define [BDB07]

$$\Theta_i = \sum_{k'} P(k'|k) n_i^{k'}, \qquad (\text{for } \ i = A, B, AB).$$

Under the un-correlation hypothesis for the degrees of neighboring nodes, i.e., $P(k'|k) = k'P(k')/\langle k\rangle$, we get the following relation for the total densities $n_i = \sum_k P(k) n_i^k$,

$$\frac{d(n_A - n_B)}{dt} = \frac{3\beta - 1}{2} \Theta_{AB}(\Theta_A - \Theta_B). \tag{2.94}$$

If we consider a small perturbation around the consensus state $n_A = 1$, with $n_A^k \gg n_B^k$ for all k, we can argue that

$$\Theta_A - \Theta_B = \sum_k k P(k)(n_A^k - n_B^k)/\langle k\rangle$$

is still positive, i.e., the consensus state is stable only for $\beta > 1/3$. In other words, the transition point does not change in heterogeneous topologies when the neutral strategy is assumed. This is in agreement with our numerical simulations, and in contrast with the other selection strategies.

For more details on negotiation dynamics and its non–equilibrium phase transitions, see [BDB07].

2.2 Elements of Haken's Synergetics

In this section we present the basics of the most powerful tool for high–dimensional chaos control, which is the *synergetics*. This powerful scientific tool to *extract order from chaos* has been developed outside of chaos theory, with intention to deal with much more complex, high–dimensional, hierarchical systems, in the realm of *synergetics*. Synergetics is an interdisciplinary field of research that was founded by H. Haken in 1969 (see [Hak83, Hak93, Hak96, Hak00]). Synergetics deals with complex systems that are composed of many individual parts (components, elements) that interact with each other and are able to produce spatial, temporal or functional structures by self–organization. In particular, synergetics searches for general principles governing self–organization irrespective of the nature of the individual parts of the systems that may belong to a variety of disciplines such as physics (lasers, fluids, plasmas), meteorology, chemistry (pattern formation by chemical reactions, including flames), biology (morphogenesis, evolution theory) movement science, brain activities, computer sciences (synergetic computer), sociology (e.g., city growth) psychology and psychiatry (including Gestalt psychology).

The aim of synergetics has been to describe processes of *spontaneous self–organization and cooperation* in complex systems built from many subsystems which themselves can be complicated nonlinear objects (like many individual neuro–muscular components of the human motion system, having their own excitation and contraction dynamics, embedded in a synergistic way to produce coordinated human movement). General properties of the subsystems are their own nonlinear/chaotic dynamics as well as mutual nonlinear/chaotic interactions. Furthermore, the systems of synergetics are *open*. The influence from outside is measured by a certain set of *control parameters* $\{\sigma\}$ (like amplitudes, frequencies and time characteristics of neuro–muscular driving forces). Processes of self-organization in synergetics, (like musculo–skeletal coordination in human motion dynamics) are observed as temporal macroscopic patterns. They are described by a small set of *order parameters* $\{o\}$, similar to those in Landau's *phase–transition theory* (named after *Nobel Laureate Lev D. Landau*) of physical systems in *thermal equilibrium* [Hak83].

Now, recall that the *measure for the degree of disorder* in any isolated, or conservative, system (such a system that does not interact with its surrounding, i.e., does neither dissipate nor gain energy) is *entropy*. The *second law of thermodynamics*[6] states that in every conservative irreversible system the entropy ever increases to its maximal value, i.e., to the total disorder of the system (or remains constant for a reversible system).

Example of such a system is conservative Hamiltonian dynamics of human skeleton in the phase–space Γ defined by all joint angles q^i and momenta p_i[7],

[6] This is the only physical law that implies the arrow of time.

[7] If we neglect joints dissipation and muscular driving forces, we are dealing with pure skeleton conservative dynamics.

defined by ordinary (conservative) Hamilton's equations

$$\dot{q}^i = \partial_{p_i} H, \qquad \dot{p}_i = -\partial_{q^i} H. \tag{2.95}$$

The basic fact of the conservative Hamiltonian system is that its phase–flow, the time evolution of equations (5.1), preserves the phase–space volume (the so–called *Liouville measure*), as proposed by the *Liouville theorem*. This might look fine at first sight, however, the preservation of phase–space volume causes *structural instability* of the conservative Hamiltonian system, i.e., the phase–space spreading effect, by which small phase regions R_t will tend to get distorted from the initial one R_0 during the system evolution. The problem is much more serious in higher dimensions than in lower dimensions, since there are so many 'directions' in which the region can locally spread. Here we see the work of the second law of thermodynamics on an irreversible process: the increase of entropy towards the total disorder/chaos [Pen89]. In this way, the conservative Hamiltonian systems of the form (5.1) cover the wide range of dynamics, from completely integrable, to completely ergodic. Biodynamics of human–like movement is probably somewhere in the middle of this range, the more DOF included in the model, the closer to the ergodic case. One can easily imagine that the conservative skeleton–like system with 300 DOF, which means 600–D system of the form (5.1), which is full of trigonometry (coming from its noncommutative rotational matrices), is probably closer to the ergodic than to the completely integrable case.

On the other hand, when we manipulate a system from the outside, by the use of certain *control parameters* $\{\sigma\}$, we can change its *degree of order* (see [Hak83, Hak93]). Consider for example *water vapor*. At elevated temperature its molecules move freely without mutual correlation. When temperature is lowered, a liquid drop is formed, the molecules now keep a mean distance between each other. Their motion is thus highly correlated. Finally, at still lower temperature, at the freezing point, water is transformed into ice crystals. The transitions between the different aggregate states, also called phases, are quite abrupt. Though the same kind of molecules are involved all the time, the macroscopic features of the three phases differ drastically.

Similar type of ordering, but not related to the thermal equilibrium conditions, occurs in *lasers*, mathematically given by Lorenz–like attractor equations. Lasers are certain types of lamps which are capable of emitting coherent light. A typical laser consists of a crystal rod filled with gas, with the following features important from the synergetics point of view: when the atoms the laser material consists of are excited or 'pumped' from the outside, they emit light waves. So, the pump power, or pump rate represents the control parameter σ. At low pump power, the waves are entirely uncorrelated as in a usual lamp. Could we hear light, it would sound like noise to us [Hak83].

When we increase the pump rate to a critical value σ_c, the noise disappears and is replaced by a pure tone. This means that the atoms emit a pure sinusoidal light wave which in turn means that the individual atoms act in a perfectly correlated way – they become self–organized. When the pump rate

is increased beyond a second critical value, the laser may periodically emit very intense and short pulses. In this way the following *instability sequence* occurs [Hak83]:

$$\text{noise} \mapsto \{\text{coherent oscillation at frequency}\,\omega_1\} \mapsto$$
$$\text{periodic pulses at frequency } \omega_2 \text{ which modulate oscillation at frequency } \omega_1$$
$$\text{i.e., no oscillation} \mapsto \text{first frequency} \mapsto \text{second frequency.}$$

Under different conditions the light emission may become *chaotic* or even *turbulent*. The frequency spectrum becomes broadened.

The laser played a crucial role in the development of synergetics for various reasons [Hak83]. In particular, it allowed detailed theoretical and experimental study of the phenomena occurring within the transition region: *lamp* $\leftrightarrow$ *laser*, where a surprising and far–reaching analogy with phase transitions of systems in thermal equilibrium was discovered. This analogy includes all basic *phase–transition effects*: a *symmetry breaking instability*, *critical slowing down* and *hysteresis effect*.

2.2.1 Phase Transitions and Synergetics

Besides water vapor, a typical example is a *ferromagnet* [Hak83]. When a ferromagnet is heated, it suddenly loses its magnetization. When temperature is lowered, the magnet suddenly regains its magnetization. What happens on a microscopic, atomic level, is this: We may visualize the magnet as being composed of many, elementary (atomic) magnets (called spins). At elevated temperature, the elementary magnets point in random directions. Their magnetic moments, when added up, cancel each other and no macroscopic magnetization results. Below a critical value of temperature T_c, the elementary magnets are lined up, giving rise to a macroscopic magnetization. Thus the *order on the microscopic level* is a cause of a *new feature* of the material *on the macroscopic level*. The change of one phase to the other one is called *phase transition*.

A thermodynamical description of a ferromagnet is based on analysis of its *free energy potential* (in thermal equilibrium conditions). The free energy $\mathcal{F}$, depends on the *control parameter* $\sigma = T$, the temperature. We seek the minimum of the potential $\mathcal{F}$ for a fixed value of magnetization o, which is called *order parameter* in Landau's theory of phase transitions (see Appendix).

This phenomenon is called a *phase transition of second order* because the second derivative (specific heat) of the free energy potential $\mathcal{F}$ is discontinuous. On the other hand, the entropy S (the first derivative of $\mathcal{F}$) itself is continuous so that this transition is also referred to as a *continuous phase transition*.

In statistical physics one also investigates the temporal change of the order parameter – magnetization o. Usually, in a more or less phenomenological manner, one assumes that o obeys an equation of the form

$$\dot{o} = -\frac{\partial \mathcal{F}}{\partial o} = -\sigma o - \beta o^3. \tag{2.96}$$

For $\sigma \to 0$ we observe a phenomenon called *critical slowing down*, because the 'particle' with coordinate o falls down the slope of the 'potential well' more and more slowly. Simple relation (2.96) is called *order parameter equation*.

We now turn to the case where the free energy potential has the form

$$\mathcal{F}(o,T) = \frac{\sigma}{2}o^2 + \frac{\gamma}{3}o^3 + \frac{\beta}{4}o^4, \tag{2.97}$$

(β and γ – positive but σ may change its sign according to $\sigma = a(T - T_c)$, $(a > 0)$). When we change the control parameter – temperature T, i.e., the parameter σ, we pass through a sequence of deformations of the potential curve.

When lowering temperature, the local minimum first remains at $o_0 = 0$. When lowering temperature, the 'particle' may fall down from o_0 to the new (global) minimum of $\mathcal{F}$ at o_1. The entropies of the two states, o_0 and o_1, differ. This phenomenon is called a *phase transition of first order* because the first derivative of the potential $\mathcal{F}$ with respect to the control parameter T is discontinuous. Since the entropy S is discontinuous this transition is also referred to as a *discontinuous phase transition*. When we now increase the temperature, is apparent that the system stays at o_1 longer than it had been before when lowering the control parameter. This represents *hysteresis effect*.

In the case of the potential (2.97) the order parameter equation gets the form

$$\dot{o} = -\sigma o - \gamma o^2 - \beta o^3.$$

Similar *disorder* $\Rightarrow$ *order* transitions occur also in various non–equilibrium systems of physics, chemistry, biology, psychology, sociology, as well as in human motion dynamics. The analogy is subsumed in Table 1.

Table 1. Phase transition analogy

System in thermal equilibrium	Non–equilibrium system
Free energy potential $\mathcal{F}$	Generalized potential V
Order parameters o_i	Order parameters o_i
$\dot{o}_i = -\frac{\partial \mathcal{F}}{\partial o_i}$	$\dot{o}_i = -\frac{\partial V}{\partial o_i}$
Temperature T	Control input u
Entropy S	System output y
Specific Heat c	System efficiency e

In the case of human motion dynamics, natural control inputs u_i are muscular torques F_i, natural system outputs y_i are joint coordinates q^i and momenta p_i, while the system efficiencies e_i represent the changes of coordinates and momenta with changes of corresponding muscular torques for the ith joint,

$$\mathrm{e}_i^q = \frac{\partial q^i}{\partial F_i}, \qquad \mathrm{e}_i^p = \frac{\partial p_i}{\partial F_i}.$$

Order parameters o_i represent certain important qualities of the human motion system, depending on muscular torques as control inputs, similar to *magnetization*, and usually defined by equations similar to (2.96) or

$$\dot{o}_i = -\sigma o - \gamma o^2 - \beta o^3,$$

with nonnegative parameters σ, β, γ, and corresponding to the second and first order phase transitions, respectively. The choice of actual order parameters is a matter of *expert knowledge* and *purpose of macroscopic system modelling* [Hak83].

2.2.2 Order Parameters in Human/Humanoid Biodynamics

Basic Hamiltonian Model of Biodynamics

To describe the biodynamics of human–like movement, namely our *covariant force law* [II05, II06a, II06b]:

$$F_i = m g_{ij} a^j, \qquad \text{that 'in plain English' reads :}$$
$$\textbf{force 1–form–field} = \textbf{mass distribution} \times \textbf{acceleration vector–field},$$

we can also start from generalized Hamiltonian vector–field X_H describing the behavior of the *human–like locomotor system*

$$\dot{q}^i = \frac{\partial H}{\partial p_i} + \frac{\partial R}{\partial p_i}, \tag{2.98}$$

$$\dot{p}_i = F_i - \frac{\partial H}{\partial q^i} + \frac{\partial R}{\partial q^i}, \tag{2.99}$$

where the vector–field X_H is generating time evolution, or *phase–flow*, of $2n$ *system variables*: n generalized coordinates (joint angles q^i) and n generalized momenta (joint angular momenta p_i), $H = H(q,p)$ represents the system's conservative energy: kinetic energy + various mechano–chemical potentials, $R = R(q,p)$ denotes the nonlinear dissipation of energy, and $F_i = F_i(t,q,p,\sigma)$ are external control forces (biochemical energy inputs). The *system parameters* include inertia tensor with mass distribution of all body segments, stiffness and damping tensors for all joints (labelled by index i, which is, for geometric reasons, written as a subscript on angle variables, and as a superscript on momentum variables), as well as amplitudes, frequencies and time characteristics of all active muscular forces (supposed to be acting in all the joints; if some of the joints are inactive, we have the affine Hamiltonian control system, see chapter 6).

The equation (2.98) is called the *velocity equation*, representing the $flow$ of the system (analogous to current in electrodynamics), while the equation (2.99) is a Newton–like *force equation*, representing the $effort$ of the system (analogous to voltage). Together, these two functions represent Hamiltonian formulation of the *biomechanical force–velocity relation* of A.V. Hill

[Hil38]. From engineering perspective, their (inner) product, $flow \cdot effort$, represents the total system's *power*, equal to the time–rate–of–change of the total system's energy (included in H, R and F_i functions). And energy itself is transformed into the *work* done by the system.

Now, the reasonably accurate musculo–skeletal biodynamics would include say a hundred DOF, which means a hundred of joint angles and a hundred of joint momenta, which further means a hundred of coupled equations of the form of (2.98–2.99). And the *full coupling* means that each angle (and momentum) includes the information of all the other angles (and momenta), the *chain coupling* means that each angle (and momentum) includes the information of all the previous (i.e., children) angles (and momenta), the *nearest neighbor coupling* includes the information of the nearest neighbors, etc.

No matter which coupling we use for modelling the dynamics of human motion, one thing is certain: the *coupling is nonlinear*. And we obviously have to fight chaos within several hundreds of variables.

Wouldn't it be better if we could somehow be able to obtain a synthetic information about the whole musculo–skeletal dynamics, synthesizing the hundreds of equations of motion of type (2.98–2.99) into a small number of equations describing the time evolution of the so–called *order parameters*? If we could do something similar to principal component analysis in multivariate statistics and neural networks, to get something like 'nonlinear factor dynamics'?

Starting from the basic system (2.98–2.99), on the lowest, *microscopic level of human movement organization*, the *order parameter equations of macroscopic synergetics* can be (at least theoretically), either exactly derived along the lines of *mezoscopic synergetics*, or phenomenologically stated by the use of the certain biophysical analogies and nonlinear identification and control techniques (a highly complex nonlinear system like human locomotor apparatus could be neither identified nor controlled by means of standard linear engineering techniques).

Mezoscopic Derivation of Order Parameters

Basic Hamiltonian equations (2.98–2.99) are in general quite complicated and can hardly be solved completely in the whole locomotor phase–space Γ, spanned by the set of possible joint vectors $\{q^i(t), p_i(t)\}$. We therefore have to restrict ourselves to local concepts for analyzing the behavior of our locomotor system. To this end we shall consider a reference musculo–skeletal state $\{q_0, p_0\}$ and its neighborhood. Following the procedures of the mezoscopic synergetics (see [Hak83, Hak93]), we assume that the reference state has the properties of an attractor and is a comparably low–dimensional object in Γ. In order to explore the behavior of our locomotor system (dependent on the set of control parameters σ) in the neighborhood of $\{q_0, p_0\}$ we look for the time development of small deviations from the reference state (to make the formalism as simple as possible, we drop the joint index in this section)

$$q(t) = q_0 + \delta q(t), \qquad p(t) = p_0 + \delta p(t),$$

and consider $\delta q(t)$ and $\delta p(t)$ as small entities. As a result we may linearize the equations of δq and δp in the vicinity of the reference state $\{q_0, p_0\}$. We get

$$\partial_t \delta q(t) = L[q_0, p_0, \sigma]\, \delta q(t), \qquad \partial_t \delta p(t) = K[q_0, p_0, \sigma]\, \delta p(t),$$

where $L[.]$ and $K[.]$ are linear matrices independent of $\delta q(t)$ and $\delta p(t)$, which can be derived from the basic Hamiltonian vector–field (2.98–2.99) by standard synergetics methods [Hak83, Hak93, Hak96, Hak00]. We now assume that we can construct a complete set of eigenvectors $\{l^{(j)}(t),\, k^{(j)}(t)\}$ corresponding to (2.96). These eigenvectors allow us to decompose arbitrary deviations $\delta q(t)$ and $\delta p(t)$ into elementary collective deviations along the directions of the eigenvectors

$$\delta q(t) = \xi_j(t)\, l^j(t), \qquad \delta p(t) = \zeta_j(t)\, k^j(t), \tag{2.100}$$

where $\xi_j(t)$ and $\zeta_j(t)$ represent the excitations of the system along the directions in the phase–space Γ prescribed by the eigenvectors $l^j(t)$ and $k^j(t)$, respectively. These amplitudes are still dependent on the set of control parameters $\{\sigma\}$. We note that the introduction of the eigenvectors $\{l^j(t),\, k^j(t)\}$ is of crucial importance. In the realm of synergetics they are considered as the *collective modes or patterns* of the system. Whereas the basic Hamiltonian equation (2.98–2.99) is formulated on the basis of the human locomotor–system variables (coordinates and momenta) of the single subsystems (joints), we can now give a new formulation which is based on these collective patterns and describes the dynamical behavior of the locomotor system in terms of these different collective patterns. Inserting relations (2.100) into the basic system (2.98–2.99) we get equations for the amplitudes $\xi_j(t)$ and $\zeta_j(t)$,

$$\dot{\xi}_i(t) = A_{ij} \cdot \xi_j(t) + \text{nonlinear terms}, \qquad \dot{\zeta}_j(t) = B_{ij} \cdot \zeta_j(t) + \text{nonlinear terms},$$

where $\cdot$ denotes the scalar product, and it is assumed that the time dependence of the linear matrices L and K is carried out by the eigenvectors leaving us with constant matrices A and B.

We now summarize the results by discussing the following time–evolution formulas for joint coordinates $q(t)$ and momenta $p(t)$,

$$q(t) = q_0 + \xi_j(t)\, l^j(t), \qquad p(t) = p_0 + \zeta_j(t)\, k^j(t), \tag{2.101}$$

which describes the time dependence of the phase vectors $q(t)$ and $p(t)$ through the evolution of the collective patterns. Obviously, the reference musculo–skeletal state $\{q_0(t), p_0(t)\}$ can be called stable when all the possible excitations $\{\xi_j(t), \zeta_j(t)\}$ decay during the curse of time. When we now change the control parameters $\{\sigma\}$ some of the $\{\xi_j(t), \zeta_j(t)\}$ can become unstable and start to grow in time. The border between decay and growth in parameter space is called a Tablecritical region. Haken has shown that the few

unstable amplitudes, denoted by u_q and u_p, change very slowly in the vicinity of a critical region, whereas the damped amplitudes, denoted by s_q and s_p, quickly decay to values which are completely prescribed by the unstable modes. This fact is expressed as the Tableslaving principle of synergetics [Hak83, Hak93, Hak96, Hak00], in our case reading as

$$s_q = s_q(u_q), \qquad s_p = s_p(u_p).$$

These relations allow us to eliminate the stable modes in (2.101), and leave us with a low–dimensional set of equations for the unstable modes which play the role of the order parameters. These Tableorder parameter equations then completely rule the behavior of our microscopic nD musculo–skeletal system on macroscopic scales near an instability.

The fundamental result of synergetics consists in the observation that on macroscopic scales new laws can be discovered which exist in their own right [Hak83]. These laws which are expressed by the order parameter equations turn out to be independent of the detailed nature of the subsystems and their interactions. As a consequence this allows us to introduce the concept of Tablenormal forms [Arn88] as a method to discus instabilities and qualitative dynamics in the neighborhood of the critical regions. This method of phenomenological synergetics allows us to start qualitative analysis from purely macroscopic considerations.

Using the so–called Tableadiabatic elimination of fast variables [Hak83], one tries to identify macroscopic quantities related to global musculo–skeletal dynamics (similar but different from the *mean–field* center–of–mass dynamics) – from experience and classifies them according to time–scale arguments. The slowest variables are usually identified with the control parameters which are assumed to be quasi static quantities. The slow macroscopic dynamics of the system has to be attributed to the order parameters. Very quickly relaxing variables have to be considered as enslaved modes.

2.2.3 Example: Synergetic Control of Biodynamics

Recall from [II05, II06a, II06b] that the basic microscopic synergetic level of human musculo–skeletal dynamics (2.98–2.99), can be viewed on the highest, *macroscopic synergetic center–of–mass organization level* of human motion dynamics as a simple *Hamilton oscillator*, physically representing the damped, sinusoidally driven pendulum (1.29) of the unit mass and length l

$$l^2\ddot{q} + \gamma\dot{q} + lg\sin q = A\cos(p_D t).$$

This equation expresses Newtonian second law of motion with the various terms on the left representing acceleration, damping, and gravitation. The angular momentum of the forcing p_D, may be different from the natural frequency of the pendulum. In order to minimize the number of adjustable parameters the equation may be rewritten in dimensionless form as

$$\ddot{q} + (1/\nu)\dot{q} + \sin q = \epsilon \cos(p_D t),$$

where ν is the damping or quality parameter, ϵ is the forcing amplitude, and p_D is the drive frequency. The low–amplitude natural angular frequency of the pendulum is unity, and time is regarded as dimensionless. This equation satisfies the necessary conditions for chaos when it is written as an extended Hamiltonian system

$$\dot{q} = p, \qquad \dot{p} = -(1/\nu)p - \sin q + \epsilon \cos \phi, \qquad \dot{\phi} = p_D. \tag{2.102}$$

The variable ϕ is introduced as the phase of the drive term. Three variables are evident and also two nonlinear coupling terms. Whether the motion is chaotic depends upon the values of the three parameters: damping, forcing amplitude and drive frequency. For some values the pendulum locks onto the driving force, oscillating in a periodic motion whose frequency is the driving frequency, possibly with some harmonics or subharmonics. But for other choices of the parameters the pendulum motion is chaotic. One may view the chaos as resulting from a subtle interplay between the tendency of the pendulum to oscillate at its 'natural' frequency and the action of the forcing term. The transitions between non–chaotic and chaotic states, due to changes in the parameters, occur in several ways and depend delicately upon the values of the parameters.

To include (in the simplest possible way) the muscle excitation–contraction dynamics, and thus make the damped, driven Hamilton oscillator (2.102) a more realistic macroscopic model for human motion dynamics, we assume that the time–dependent forcing amplitude $\epsilon = \epsilon(t)$ has the form of a low pass filter, a characteristic feature of biological systems, given by first–order transfer function $\frac{K}{Ts+1}$. Here K denotes gain of the filter and T its time constant.

Therefore, macroscopic mechanical model of human motion dynamics gets the fully–functional form

$$\ddot{q} + (1/\nu)\dot{q} + \sin q = K(1 - \mathrm{e}^{-t/T}) \cos(p_D t),$$

which can be rewritten in the form of *extended Hamilton oscillator*

$$\dot{q} = p, \qquad \dot{p} = -(1/\nu)p - \sin q + K(1 - \mathrm{e}^{-t/T}) \cos \phi, \qquad \dot{\phi} = p_D. \tag{2.103}$$

Now, to effectively control the macroscopic HML model (2.103), we can use two standard nonlinear–control techniques:

A. Adaptive Lie–derivative based geometric control; and
B. Adaptive fuzzy–logic based AI control.

2.2.4 Example: Chaotic Psychodynamics of Perception

Perceptual alternation phenomena of ambiguous figures have been studied for a long time. Figure–ground, perspective (depth) and semantic ambiguities are

well known (see, e.g., [Att71, Hak91]). When we view the *Necker cube*, which is a classic example of perspective alternation, a part of the figure is perceived either as front or back of a cube and our perception switches between the two different interpretations (see Figure 2.7). In this circumstance the external stimulus is kept constant, but perception undergoes involuntary and random–like change. The measurements have been quantified in psychophysical experiments and it becomes evident that the times between such changes are approximately Gamma distributed [BMA72, BCR82, Hak91].

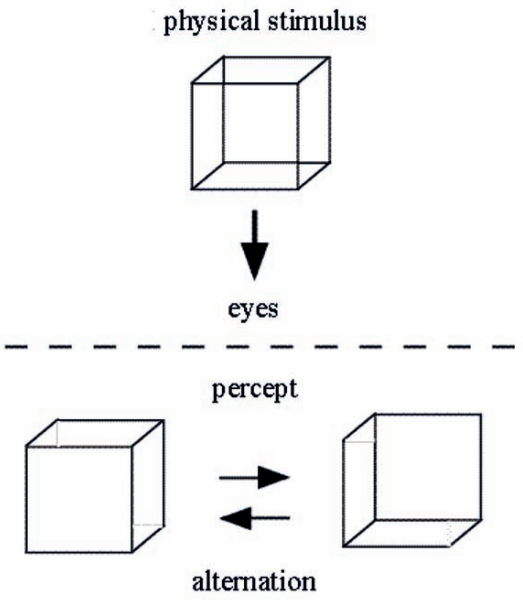

Fig. 2.7. Perception of the Necker cube with its two alternative interpretations (modified and adapted from [NNM00]).

Mathematical model approaches to explaining the facts have been made mainly from three situations based on the *synergetics* [DT89, DT90, CA93], the BSB (brain–state–in–a–box) neural network model [KA85, RMS90, MM95], and the PDP (parallel distributed processing) schema model [RM86, SKW95, IN96]. Common to these approaches is that top–down designs are applied so that the model can be manipulable by a few parameters and upon this basis fluctuating sources are brought in. The major interests seem to be not in the relation between the whole function and its element (neuron), but in the model building at the phenomenological level.

So far diverse types of chaotic dynamics have been confirmed at several hierarchical levels in the real neural systems from single cells to cortical networks (e.g. ionic channels, spike trains from cells, EEG) [Arb95].

Following [NNM00], in this section we present a perception model of ambiguous patterns based on the chaotic neural network from the viewpoint of bottom–up approach [NNM97], aiming at the functioning of chaos in dynamic perceptual processes.

The chaotic neural network (CNN) composed of N chaotic neurons is described as [ATT90, NKF97] (summation upon repeated indices is always understood)

$$X_i(t+1) = f(\eta_i(t+1) + \zeta_i(t+1)), \tag{2.104}$$

$$\eta_i(t+1) = w_{ij} \sum_{d=0}^{t} k_f^d X_j(t-d), \tag{2.105}$$

$$\zeta_i(t+1) = -\alpha \sum_{d=0}^{t} k_r^d X_i(t-d) - \theta_i, \tag{2.106}$$

where X_i : output of neuron $i(-1 \leq X_i \leq 1)$, w_{ij} : synaptic weight from neuron j to neuron i, θ_i : threshold of neuron i, $k_f(k_r)$: decay factor for the feedback (refractoriness) $(0 \leq k_f,\ k_r < 1)$, α : refractory scaling parameter, f : output function defined by $f(y) = \tanh(y/2\varepsilon)$ with the steepness parameter ε. Owing to the exponentially decaying form of the past influence, (2.105) and (2.106) can be reduced to

$$\eta_i(t+1) = k_f \eta_i(t) + w_{ij} X_j(t), \tag{2.107}$$

$$\zeta_i(t+1) = k_r \zeta_i(t) - \alpha X_i(t) + a, \tag{2.108}$$

where a is temporally constant $a \equiv -\theta_i(1-k_r)$. All neurons are updated in parallel, that is, synchronously. The network corresponds to the conventional *Hopfield discrete–time network*:

$$X_i(t+1) = f\left[w_{ij} X_j(t) - \theta_i\right], \tag{2.109}$$

when $\alpha = k_f = k_r = 0$ (Hopfield network point (HNP)). The asymptotical stability and chaos in discrete–time neural networks are theoretically investigated in [MW89, CA97]. The stochastic fluctuation $\{F_i\}$ is attached to (2.109) of HNP together with the external stimulus $\{\sigma_i\}$:

$$X_i(t+1) = f\left[w_{ij} X_j(t) + \sigma_i + F_i(t)\right],$$
$$\text{where} \quad \begin{cases} < F_i(t) >= 0 \\ < F_i(t) F_j(t') >= D^2 \delta_{tt'} \delta_{ij}. \end{cases}$$

Under external stimuli, (2.104) is influenced as

$$X_i(t+1) = f\left[\eta_i(t+1) + \zeta_i(t+1) + \sigma_i\right], \tag{2.110}$$

where $\{\sigma_i\}$ is the effective term by external stimuli. This is a simple and un–artificial incorporation of stimuli as the changes of neural active potentials.

The two competitive interpretations are embedded in the network as minima of the energy map (see Figure 2.8):

$$E = -\frac{1}{2} w_{ij} X_i X_j,$$

at HNP. This is done by using a iterative perception learning rule for $p(< N)$ patterns $\{\xi_i^\mu\} \equiv (\xi_1^\mu, \cdots, \xi_N^\mu), (\mu = 1, \cdots, p; \xi_i^\mu = \ +1 or -1)$ in the form :

$$w_{ij}^{new} = w_{ij}^{old} + \sum_\mu \delta w_{ij}^\mu, \qquad \text{with} \qquad \delta w_{ij}^\mu = \frac{1}{N} \theta(1 - \gamma_i^\mu) \xi_i^\mu \xi_j^\mu,$$
$$\text{where} \qquad \gamma_i^\mu \equiv \xi_i^\mu w_{ij} \xi_j^\mu,$$

and $\theta(h)$ is the unit step function. The learning mode is separated from the performance mode by (2.110).

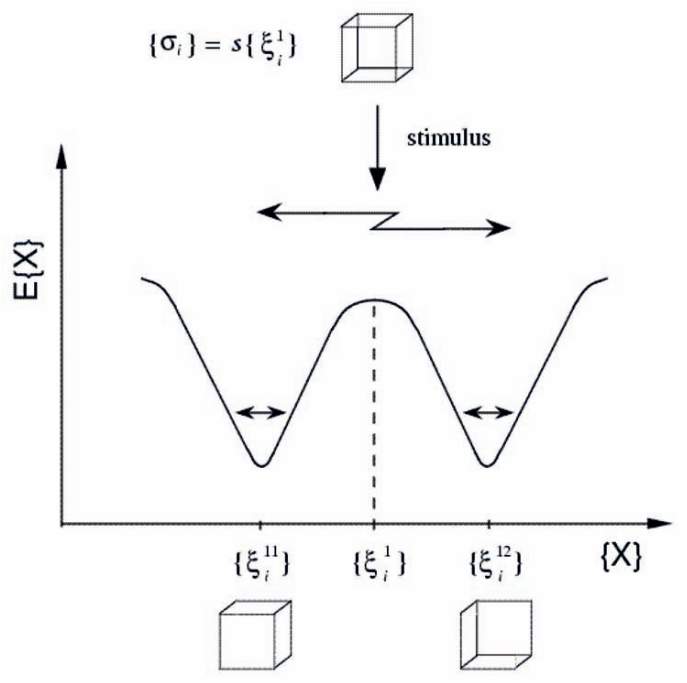

Fig. 2.8. Conceptual psychodynamic model of [NNM00], illustrating state transitions induced by chaotic activity.

Simulations of the CNN have shown that the neural chaos leads to perceptual alternations as responses to ambiguous stimuli in the chaotic neural network. Its emergence is based on the simple process in a realistic bottom–up framework. In the same stage, similar results can not be obtained by

the stochastic activity. This simulation suggests functional usefulness of the chaotic activity in perceptual systems even at higher cognitive levels. The perceptual alternation appears to be an inherent feature built in the chaotic neuron assembly. It may be interesting to study the brain with the experimental technique (e.g., functional MRI) under the circumstance where the perceptual alternation is running [NNM00].

2.2.5 Kick Dynamics and Dissipation–Fluctuation Theorem

Deterministic Delayed Kicks

Following [Hak02], we consider the mechanical example of a soccer ball that is kicked by a soccer player and rolls over grass, whereby its motion will be slowed down. We start with the Newton's (second) law of motion, $m\dot{v} = force$, and in order to get rid of superfluous constants, we put temporarily $m = 1$. The $force$ on the r.h.s. consists of the damping force -$\gamma v(t)$ of the grass (where γ is the damping constant) and the sharp force $F(t) = s\delta(t - \sigma)$ of the individual kick occurring at time $t = \sigma$ (where s is the strength of the kick, and δ is the Dirac's 'delta' function). In this way, the (single) *kick equation* of the ball motion becomes

$$\dot{v} = -\gamma v(t) + s\delta(t - \sigma), \tag{2.111}$$

with the general solution

$$v(t) = sG(t - \sigma),$$

where $G(t - \sigma)$ is the *Green's function*[8]

$$G(t - \sigma) = \begin{cases} 0 & \text{for } t < \sigma \\ \mathrm{e}^{-\gamma(t-\sigma)} & \text{for } t \geq \sigma \end{cases}.$$

Now, we can generalize the above to N kicks with individual strengths s_j, occurring at a sequence of times $\{\sigma_j\}$, so that the total kicking force becomes

$$F(t) = \sum_{j=1}^{N} s_j \delta(t - \sigma_j).$$

In this way, we get the *multi–kick equation* of the ball motion

[8] This is the Green's function of the first order system (2.111). Similarly, the Green's function

$$G(t - \sigma) = \begin{cases} 0 & \text{for } t < \sigma \\ (t - \sigma)\mathrm{e}^{-\gamma(t-\sigma)} & \text{for } t \geq \sigma \end{cases}$$

corresponds to the second order system

$$\left(\frac{d}{dt} + \gamma\right)^2 G(t - \sigma) = \delta(t - \sigma).$$

$$\dot{v} = -\gamma v(t) + \sum_{j=1}^{N} s_j \delta(t - \sigma_j),$$

with the general solution

$$v(t) = \sum_{j=1}^{N} s_j G(t - \sigma_j). \tag{2.112}$$

As a final generalization, we would imagine that the kicks are continuously exerted on the ball, so that kicking force becomes

$$F(t) = \int_{t_0}^{T} s(\sigma)\delta(t - \sigma)d\sigma \equiv \int_{t_0}^{T} d\sigma F(\sigma)\delta(t - \sigma),$$

so that the continuous multi–kick equation of the ball motion becomes

$$\dot{v} = -\gamma v(t) + \int_{t_0}^{T} s(\sigma)\delta(t - \sigma)d\sigma \equiv -\gamma v(t) + \int_{t_0}^{T} d\sigma F(\sigma)\delta(t - \sigma),$$

with the general solution

$$v(t) = \int_{t_0}^{T} d\sigma F(\sigma)G(t - \sigma) = \int_{t_0}^{T} d\sigma F(\sigma)\mathrm{e}^{-\gamma(t-\sigma)}. \tag{2.113}$$

Random Kicks and Langevin Equation

We now denote the times at which kicks occur by t_j and indicate their direction in a one–dimensional game by $(\pm 1)_j$, where the choice of the plus or minus sign is random (e.g., throwing a coin). Thus the kicking force can be written in the form [Hak02]

$$F(t) = s \sum_{j=1}^{N} \delta(t - t_j)(\pm 1)_j, \tag{2.114}$$

where for simplicity we assume that all kicks have the same strength s. When we observe many games, then we may perform an average $< ... >$ over all these different *performances*,

$$< F(t) >= s < \sum_{j=1}^{N} \delta(t - t_j)(\pm 1)_j > . \tag{2.115}$$

Since the direction of the kicks is assumed to be independent of the time at which the kicks happen, we may split (2.115) into the product

$$< F(t) >= s < \sum_{j=1}^{N} \delta(t - t_j) >< (\pm 1)_j > .$$

As the kicks are assumed to happen with equal frequency in both directions, we get the cancellation

$$< (\pm 1)_j >= 0,$$

which implies that the average kicking force also vanishes,

$$< F(t) >= 0.$$

In order to characterize the strength of the force (2.114), we consider a quadratic expression in F, e.g., by calculating the *correlation function* for two times t, t',

$$< F(t)F(t') >= s^2 < \sum_j \delta(t - t_j)(\pm 1)_j \sum_k \delta(t' - t_k)(\pm 1)_k > .$$

As the ones for $j \neq k$ will cancel each other and for $j = k$ will become 1, the correlation function becomes a single sum

$$< F(t)F(t') >= s^2 < \sum_j \delta(t - t_j)\delta(t' - t_k) >, \tag{2.116}$$

which is usually evaluated by assuming the *Poisson process* for the times of the kicks.

Now, proper description of random motion is given by *Langevin rate equation*, which describes the *Brownian motion*: when a particle is immersed in a fluid, the velocity of this particle is slowed down by a force proportional to its velocity and the particle undergoes a zig–zag motion (the particle is steadily pushed by much smaller particles of the liquid in a random way). In physical terminology, we deal with the behavior of a system (particle) which is coupled to a *heat bath* or reservoir (namely the liquid). The heat bath has two effects [Hak02]:

A. It decelerates the mean motion of the particle; and
B. It causes statistical fluctuation.

The standard Langevin equation has the form

$$\dot{v} = -\gamma v(t) + F(t), \tag{2.117}$$

where $F(t)$ is a *fluctuating force* with the following properties:

A. Its statistical average (2.115) vanishes; and
B. Its correlation function (2.116) is given by

$$< F(t)F(t') >= Q\delta(t - t_0), \tag{2.118}$$

where $t_0 = T/N$ denotes the mean free time between kicks, and $Q = s^2/t_0$ is the *random fluctuation*.

The general solution of the Langevin equation (2.117) is given by (2.113).

The average velocity vanishes, $< v(t) >= 0$, as both directions are possible and cancel each other. Using the integral solution (2.113) we get

$$< v(t)v(t') >=< \int_{t_0}^{t} d\sigma \int_{t_0}^{t'} d\sigma' F(\sigma)F(\sigma')\mathrm{e}^{-\gamma(t-\sigma)}\mathrm{e}^{-\gamma(t'-\sigma')} >,$$

which, in the steady–state, reduces to

$$< v(t)v(t') >= \frac{Q}{2\gamma}\mathrm{e}^{-\gamma(t-\sigma)},$$

and for equal times

$$< v(t)^2 >= \frac{Q}{2\gamma}.$$

If we now repeat all the steps performed so far with $m \neq 1$, the final result reads

$$< v(t)^2 >= \frac{Q}{2\gamma m}. \tag{2.119}$$

Now, according to thermodynamics, the *mean kinetic energy* of a particle is given by

$$\frac{m}{2} < v(t)^2 >= \frac{1}{2}k_B T, \tag{2.120}$$

where T is the (absolute) temperature, and k_B is the Boltzman's constant. Comparing (2.119) and (2.120), we get the important Einstein's result

$$Q = 2\gamma k_B T,$$

which says that whenever there is damping, i.e., $\gamma \neq 0$, then there are random fluctuations (or noise) Q. In other words, fluctuations or noise are inevitable in any physical system. For example, in a resistor (with the resistance R) the electric field E fluctuates with a correlation function (similar to (2.118))

$$< E(t)E(t') >= 2Rk_B T\delta(t - t_0).$$

This is the simplest example of the *fluctuation–dissipation theorem*.

2.3 Synergetics of Recurrent and Attractor Neural Networks

Recall that *recurrent neural networks* are neural networks with synaptic feedback loops. Provided that we restrict ourselves to large neural systems, we can apply to their analysis tools from statistical mechanics. Here, we have two possibilities. Under the common conditions of synaptic symmetry, the stochastic process of evolving neuron states leads towards an equilibrium situation

where the microscopic state probabilities are known, where the classical techniques of *equilibrium statistical mechanics* can be applied. On the other hand, for non–symmetric networks, where the asymptotic (stationary) statistics are not known, synergetic techniques from *non–equilibrium statistical mechanics* are the only tools available for analysis. Here, the 'natural' set of macroscopic *order parameters* to be calculated can be defined in practice as the smallest set which will obey closed deterministic equations in the limit of an infinitely large network.

Being high–dimensional nonlinear systems with extensive feedback, the dynamics of recurrent neural networks are generally dominated by a wealth of different attractors, and the practical use of recurrent neural networks (in both biology and engineering) lies in the potential for creation and manipulation of these attractors through adaptation of the network parameters (synapses and thresholds). Input fed into a recurrent neural network usually serves to induce a specific initial configuration (or firing pattern) of the neurons, which serves as a cue, and the 'output' is given by the (static or dynamic) attractor which has been triggered by this cue. The most familiar types of recurrent neural network models, where the idea of creating and manipulating attractors has been worked out and applied explicitly, are the so–called *attractor neural networks* for associative memory, designed to store and retrieve information in the form of neuronal firing patterns and/or sequences of neuronal firing patterns. Each pattern to be stored is represented as a microscopic state vector. One then constructs synapses and thresholds such that the dominant attractors of the network are precisely the pattern vectors (in the case of static recall), or where, alternatively, they are trajectories in which the patterns are successively generated microscopic system states. From an initial configuration (the 'cue', or input pattern to be recognized) the system is allowed to evolve in time autonomously, and the final state (or trajectory) reached can be interpreted as the pattern (or pattern sequence) recognized by network from the input. For such programmes to work one clearly needs recurrent neural networks with extensive *ergodicity breaking*: the state vector will during the course of the dynamics (at least on finite time–scales) have to be confined to a restricted region of state space (an 'ergodic component'), the location of which is to depend strongly on the initial conditions. Hence our interest will mainly be in systems with many attractors. This, in turn, has implications at a theoretical/mathematical level: solving models of recurrent neural networks with extensively many attractors requires advanced tools from disordered systems theory, such as statical replica theory and dynamical partition function analysis.

The *equilibrium* statistical mechanical techniques can provide much detailed quantitative information on the behavior of recurrent neural networks, but they obviously have serious restrictions. The first one is that, by definition, they will only provide information on network properties in the stationary state. For associative memories, for instance, it is not clear how one can calculate quantities like sizes of domains of attraction without solving the

dynamics. The second, and more serious, restriction is that for equilibrium statistical mechanics to apply the dynamics of the network under study must obey detailed balance, i.e., absence of microscopic probability currents in the stationary state. For recurrent networks in which the dynamics take the form of a stochastic alignment of neuronal firing rates to post–synaptic potentials which, in turn, depend linearly on the firing rates, this requirement of detailed balance usually implies symmetry of the synaptic matrix. From a physiological point of view this requirement is clearly unacceptable, since it is violated in any network that obeys Dale's law as soon as an excitatory neuron is connected to an inhibitory one. Worse still, in any network of graded–response neurons detailed balance will always be violated, even when the synapses are symmetric. The situation will become even worse when we turn to networks of yet more realistic (spike–based) neurons, such as integrate-and-fire ones. In contrast to this, *non–equilibrium* statistical mechanical techniques, it will turn out, do not impose such biologically non–realistic restrictions on neuron types and synaptic symmetry, and they are consequently the more appropriate avenue for future theoretical research aimed at solving biologically more realistic models (for details, see [Coo01, SC00, SC01, CKS05]).

2.3.1 Stochastic Dynamics of Neuronal Firing States

Recall that the simplest non–trivial definition of a recurrent neural network is that where N binary neurons $\sigma_i \in \{-1, 1\}$ (in which the states '1' and '-1' represent firing and rest, respectively) respond iteratively and synchronously to post–synaptic potentials (or local fields) $h_i(\boldsymbol{\sigma})$, with $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_N)$. The fields are assumed to depend linearly on the instantaneous neuron states (summation convention upon repeated indices is always used):

$$\text{Parallel Dynamics:} \quad \sigma_i(\ell+1) = \operatorname{sgn}\left[h_i(\boldsymbol{\sigma}(\ell)) + T\eta_i(\ell)\right], \quad h_i(\sigma) = J_{ij}\sigma_j + \theta_i. \tag{2.121}$$

The stochasticity is in the independent random numbers $\eta_i(\ell) \in \mathbb{R}$ (representing threshold noise), which are all drawn according to some distribution $w(\eta)$. The parameter T is introduced to control the amount of noise. For $T = 0$ the process (2.121) is deterministic: $\sigma_i(\ell + 1) = \operatorname{sgn}[h_i(\sigma(\ell))]$. The opposite extreme is choosing $T = \infty$, here the system evolution is fully random. The external fields θ_i represent neural thresholds and/or external stimuli, J_{ij} represents the synaptic efficacy at the junction $j \to i$ ($J_{ij} > 0$ implies excitation, $J_{ij} < 0$ inhibition). Alternatively we could decide that at each iteration step ℓ only a single randomly drawn neuron σ_{i_ℓ} is to undergo an update of the type (2.121):

$$\text{Sequential Dynamics:} \quad \begin{array}{l} i \neq i_\ell : \sigma_i(\ell+1) = \sigma_i(\ell), \\ i = i_\ell : \sigma_i(\ell+1) = \operatorname{sgn}\left[h_i(\boldsymbol{\sigma}(\ell)) + T\eta_i(\ell)\right], \end{array} \tag{2.122}$$

with the local fields as in (2.121). The stochasticity is now both in the independent random numbers $\eta_i(\ell)$ (the threshold noise) and in the site i_ℓ to be

updated, drawn randomly from the set $\{1, \ldots, N\}$. For simplicity we assume $w(-\eta) = w(\eta)$, and define

$$g(z) = 2\int_0^z d\eta\; w(\eta): \qquad g(-z) = -g(z), \qquad \lim_{z\to\pm\infty} g(z) = \pm 1, \qquad \partial_z g(z) \geq 0.$$

Popular choices for the threshold noise distributions are

$$w(\eta) = (2\pi)^{-\frac{1}{2}} \mathrm{e}^{-\frac{1}{2}\eta^2}: \qquad g(z) = \mathrm{Erf}[z/\sqrt{2}],$$
$$w(\eta) = \frac{1}{2}[1 - \tanh^2(\eta)]: \qquad g(z) = \tanh(z).$$

Now, from the microscopic equations (2.121,2.122), which are suitable for numerical simulations, we can derive an equivalent but mathematically more convenient description in terms of microscopic state probabilities $p_\ell(\boldsymbol{\sigma})$. Equations (2.121,2.122) state that, if the system state $\boldsymbol{\sigma}(\ell)$ is given, a neuron i to be updated will obey

$$\mathrm{Prob}\,[\sigma_i(\ell+1)] = \frac{1}{2}\,[1 + \sigma_i(\ell+1)\; g[\beta h_i(\boldsymbol{\sigma}(\ell))]]\,, \tag{2.123}$$

with $\beta = T^{-1}$. In the case (2.121) this rule applies to all neurons, and thus we simply get $p_{\ell+1}(\boldsymbol{\sigma}) = \prod_{i=1}^N \frac{1}{2}\,[1 + \sigma_i\; g[\beta h_i(\boldsymbol{\sigma}(\ell))]]$. If, on the other hand, instead of $\boldsymbol{\sigma}(\ell)$ only the probability distribution $p_\ell(\boldsymbol{\sigma})$ is given, this expression for $p_{\ell+1}(\boldsymbol{\sigma})$ is to be averaged over the possible states at time ℓ:

$$\text{Parallel Dynamics}: \qquad p_{\ell+1}(\boldsymbol{\sigma}) = \sum_{\boldsymbol{\sigma}'} W\,[\boldsymbol{\sigma};\boldsymbol{\sigma}']\,p_\ell(\boldsymbol{\sigma}'), \tag{2.124}$$
$$W\,[\boldsymbol{\sigma};\boldsymbol{\sigma}'] = \prod_{i=1}^N \frac{1}{2}\,[1 + \sigma_i\; g[\beta h_i(\boldsymbol{\sigma}')]]\,.$$

This is the standard representation of a *Markov chain*. Also the sequential process (2.122) can be formulated in terms of probabilities, but here expression (2.123) applies only to the randomly drawn candidate i_ℓ. After averaging over all possible realisations of the sites i_ℓ we get

$$p_{\ell+1}(\boldsymbol{\sigma}) = \frac{1}{N}\left\{[\prod_{j\neq i}\delta_{\sigma_j,\sigma_j(\ell)}]\;\frac{1}{2}\,[1 + \sigma_i\; g[\beta h_i(\boldsymbol{\sigma}(\ell))]]\right\},$$

with the Kronecker symbol: $\delta_{ij} = 1$, if $i = j$ and $\delta_{ij} = 0$, otherwise. If, instead of $\boldsymbol{\sigma}(\ell)$, the probabilities $p_\ell(\boldsymbol{\sigma})$ are given, this expression is to be averaged over the possible states at time ℓ, with the result:

$$p_{\ell+1}(\boldsymbol{\sigma}) = \frac{1}{2N}\,[1 + \sigma_i\; g[\beta h_i(\boldsymbol{\sigma})]]\,p_\ell(\boldsymbol{\sigma})$$
$$+\;\frac{1}{2N}\,[1 + \sigma_i\; g[\beta h_i(F_i\boldsymbol{\sigma})]]\,p_\ell(F_i\boldsymbol{\sigma}),$$

with the state-flip operators $F_i\Phi(\boldsymbol{\sigma}) = \Phi(\sigma_1, \ldots, \sigma_{i-1}, -\sigma_i, \sigma_{i+1}, \ldots, \sigma_N)$. This equation can again be written in the standard form

$$p_{\ell+1}(\boldsymbol{\sigma}) = \sum_{\boldsymbol{\sigma}'} W\left[\boldsymbol{\sigma}; \boldsymbol{\sigma}'\right] p_\ell(\boldsymbol{\sigma}'),$$

but now with the transition matrix

$$\text{Sequential Dynamics:} \quad W\left[\boldsymbol{\sigma}; \boldsymbol{\sigma}'\right] = \delta_{\boldsymbol{\sigma},\boldsymbol{\sigma}'} + \frac{1}{N}\left\{w_i(F_i\boldsymbol{\sigma})\delta_{\boldsymbol{\sigma},F_i\boldsymbol{\sigma}'} - w_i(\boldsymbol{\sigma})\delta_{\boldsymbol{\sigma},\boldsymbol{\sigma}'}\right\}, \tag{2.125}$$

$$\text{where} \qquad \delta_{\boldsymbol{\sigma},\boldsymbol{\sigma}'} = \prod_i \delta_{\sigma_i,\sigma_i'} \qquad \text{and} \qquad w_i(\boldsymbol{\sigma}) = \frac{1}{2}\left[1 - \sigma_i \tanh\left[\beta h_i(\boldsymbol{\sigma})\right]\right]. \tag{2.126}$$

Note that, as soon as $T > 0$, the two transition matrices $W[\boldsymbol{\sigma}; \boldsymbol{\sigma}']$ in (2.124,2.125) both describe *ergodic systems*: from any initial state $\boldsymbol{\sigma}'$ one can reach any final state $\boldsymbol{\sigma}$ with nonzero probability in a finite number of steps (being one in the parallel case, and N in the sequential case). It now follows from the standard theory of stochastic processes (see e.g., [Kam92, Gar85]) that in both cases the system evolves towards a unique stationary distribution $p_\infty(\boldsymbol{\sigma})$, where all probabilities $p_\infty(\boldsymbol{\sigma})$ are non–zero [Coo01, SC00, SC01, CKS05].

The above processes have the (mathematically and biologically) less appealing property that time is measured in discrete units. For the sequential case we will now assume that the *duration* of each of the iteration steps is a continuous random number (for parallel dynamics this would make little sense, since all updates would still be made in full synchrony). The statistics of the durations are described by a function $\pi_\ell(t)$, defined as the probability that at time t precisely ℓ updates have been made. Upon denoting the previous discrete-time probabilities as $\hat{p}_\ell(\boldsymbol{\sigma})$, our new process (which now includes the randomness in step duration) will be described by

$$p_t(\boldsymbol{\sigma}) = \sum_{\ell \geq 0} \pi_\ell(t)\hat{p}_\ell(\boldsymbol{\sigma}) = \sum_{\ell \geq 0} \pi_\ell(t) \sum_{\boldsymbol{\sigma}'} W^\ell\left[\boldsymbol{\sigma}; \boldsymbol{\sigma}'\right] p_0(\boldsymbol{\sigma}'),$$

and time has become a continuous variable. For $\pi_\ell(t)$ we make the Poisson choice,

$$\pi_\ell(t) = \frac{1}{\ell+}\left(\frac{t}{\Delta}\right)^\ell \mathrm{e}^{-t/\Delta}.$$

From $\langle \ell \rangle_\pi = t/\Delta$ and $\langle \ell^2 \rangle_\pi = t/\Delta + t^2/\Delta^2$ it follows that Δ is the average duration of an iteration step, and that the relative deviation in ℓ at a given t vanishes for $\Delta \to 0$ as $\sqrt{\langle \ell^2 \rangle_\pi - \langle \ell \rangle_\pi^2}/\langle \ell \rangle_\pi = \sqrt{\Delta/t}$. The nice properties of the Poisson distribution under temporal derivation allow us to derive:

$$\Delta \dot{p}_t(\boldsymbol{\sigma}) = \sum_{\boldsymbol{\sigma}'} W\left[\boldsymbol{\sigma}; \boldsymbol{\sigma}'\right] p_t(\boldsymbol{\sigma}') - p_t(\boldsymbol{\sigma}).$$

For sequential dynamics, we choose $\Delta = \frac{1}{N}$ so that, as in the parallel case, in one time unit each neuron will on average be updated once. The master

equation corresponding to (2.125) acquires the form $w_i(\boldsymbol{\sigma})$; (2.126) now play the role of *transition rates*. The choice $\Delta = \frac{1}{N}$ implies $\sqrt{\langle \ell^2 \rangle_\pi - \langle \ell \rangle_\pi^2}/\langle \ell \rangle_\pi = \sqrt{1/Nt}$, so we will still for $N \to \infty$ no longer have uncertainty in where we are on the t axis.

Alternatively, we could start with continuous neuronal variables σ_i (representing e.g., firing frequencies or oscillator phases), where $i = 1, \ldots, N$, and with stochastic equations of the form

$$\sigma_i(t + \Delta) = \sigma_i(t) + \Delta f_i(\boldsymbol{\sigma}(t)) + \sqrt{2T\Delta}\xi_i(t). \tag{2.127}$$

Here, we have introduced (as yet unspecified) deterministic state–dependent forces $f_i(\boldsymbol{\sigma})$, and uncorrelated Gaussian distributed random forces $\xi_i(t)$ (the noise), with

$$\langle \xi_i(t) \rangle = 0 \qquad \text{and} \qquad \langle \xi_i(t)\xi_j(t') \rangle = \delta_{ij}\delta_{t,t'}.$$

As before, the parameter T controls the amount of noise in the system, ranging from $T = 0$ (deterministic dynamics) to $T = \infty$ (completely random dynamics). If we take the limit $\Delta \to 0$ in (2.127) we find a Langevin equation (with a continuous time variable) [Coo01, SC00, SC01, CKS05]:

$$\dot{\sigma}_i(t) = f_i(\boldsymbol{\sigma}(t)) + \eta_i(t). \tag{2.128}$$

This equation acquires its meaning only as the limit $\Delta \to 0$ of (2.127). The moments of the new noise variables $\eta_i(t) = \xi_i(t)\sqrt{2T/\Delta}$ in (2.128) are given by

$$\langle \eta_i(t) \rangle = 0 \qquad \text{and} \qquad \langle \eta_i(t)\eta_j(t') \rangle = 2T\delta_{ij}\delta(t - t').$$

This can be derived from the moments of the $\xi_i(t)$. For instance:

$$\begin{aligned} \langle \eta_i(t)\eta_j(t') \rangle &= \lim_{\Delta \to 0} \frac{2T}{\Delta} \langle \xi_i(t)\xi_j(t') \rangle \\ &= 2T\delta_{ij} \lim_{\Delta \to 0} \frac{1}{\Delta} \delta_{t,t'} = 2TC\delta_{ij}\delta(t - t'). \end{aligned}$$

The constant C is found by summing over t', before taking the limit $\Delta \to 0$, in the above equation:

$$\begin{aligned} \int dt' \, \langle \eta_i(t)\eta_j(t') \rangle &= \lim_{\Delta \to 0} 2T \sum_{t'=-\infty}^{\infty} \langle \xi_i(t)\xi_j(t') \rangle \\ &= 2T\delta_{ij} \lim_{\Delta \to 0} \sum_{t'=-\infty}^{\infty} \delta_{t,t'} = 2T\delta_{ij}. \end{aligned}$$

Thus $C = 1$, which indeed implies $\langle \eta_i(t)\eta_j(t') \rangle = 2T\delta_{ij}\delta(t - t')$. More directly, one can also calculate the moment partition function:

$$\langle \mathrm{e}^{i\int dt\psi_i(t)\eta_i(t)}\rangle = \lim_{\Delta\to 0}\prod_{i,t}\int \frac{dz}{\sqrt{2\pi}}\,\mathrm{e}^{-\frac{1}{2}z^2+iz\psi_i(t)\sqrt{2T\Delta}} \tag{2.129}$$

$$= \lim_{\Delta\to 0}\prod_{i,t}\mathrm{e}^{-T\Delta\psi_i^2(t)} = \mathrm{e}^{-T\int dt\sum_i\psi_i^2(t)}. \tag{2.130}$$

On the other hand, a mathematically more convenient description of the process (2.128) is provided by the Fokker–Planck equation for the microscopic state probability density $p_t(\boldsymbol{\sigma}) = \langle\delta[\boldsymbol{\sigma}-\boldsymbol{\sigma}(t)]\rangle$, which we will now derive. For the discrete–time process (2.127) we expand the $\delta-$distribution in the definition of $p_{t+\Delta}(\boldsymbol{\sigma})$ (in a distributional sense) [Coo01, SC00, SC01, CKS05]:

$$\begin{aligned} p_{t+\Delta}(\boldsymbol{\sigma}) - p_t(\boldsymbol{\sigma}) &= \langle\delta\left[\boldsymbol{\sigma}-\boldsymbol{\sigma}(t)-\Delta\mathbf{f}(\boldsymbol{\sigma}(t))-\sqrt{2T\Delta}\boldsymbol{\xi}(t)\right]\rangle - \langle\delta[\boldsymbol{\sigma}-\boldsymbol{\sigma}(t)]\rangle \\ &= -\frac{\partial}{\partial\sigma_i}\langle\delta[\boldsymbol{\sigma}-\boldsymbol{\sigma}(t)]\left[\Delta f_i(\boldsymbol{\sigma}(t))+\sqrt{2T\Delta}\xi_i(t)\right]\rangle \\ &\quad + T\Delta\frac{\partial^2}{\partial\sigma_i\partial\sigma_j}\langle\delta[\boldsymbol{\sigma}-\boldsymbol{\sigma}(t)]\xi_i(t)\xi_j(t)\rangle + \mathcal{O}(\Delta^{\frac{3}{2}}). \end{aligned}$$

The variables $\boldsymbol{\sigma}(t)$ depend only on noise variables $\xi_j(t')$ with $t' < t$, so that for any function A,

$$\begin{aligned} \langle A[\boldsymbol{\sigma}(t)]\xi_i(t)\rangle &= \langle A[\boldsymbol{\sigma}(t)]\rangle\langle\xi_i(t)\rangle = 0, \qquad \text{and} \\ \langle A[\boldsymbol{\sigma}(t)]\xi_i(t)\xi_j(t)\rangle &= \delta_{ij}\langle A[\boldsymbol{\sigma}(t)]\rangle. \end{aligned}$$

As a consequence, we have:

$$\begin{aligned} \frac{1}{\Delta}\left[p_{t+\Delta}(\boldsymbol{\sigma})-p_t(\boldsymbol{\sigma})\right] &= -\frac{\partial}{\partial\sigma_i}\langle\delta[\boldsymbol{\sigma}-\boldsymbol{\sigma}(t)]f_i(\boldsymbol{\sigma}(t))\rangle \\ &\quad + T\frac{\partial^2}{\partial\sigma_i^2}\langle\delta[\boldsymbol{\sigma}-\boldsymbol{\sigma}(t)]\rangle + \mathcal{O}(\Delta^{\frac{1}{2}}) \\ &= -\frac{\partial}{\partial\sigma_i}\left[p_t(\boldsymbol{\sigma})f_i(\boldsymbol{\sigma})\right] + T\frac{\partial^2}{\partial\sigma_i^2}p_t(\boldsymbol{\sigma}) + \mathcal{O}(\Delta^{\frac{1}{2}}). \end{aligned}$$

By taking the limit $\Delta\to 0$ we then arrive at the Fokker–Planck equation:

$$\dot{p}_t(\boldsymbol{\sigma}) = -\frac{\partial}{\partial\sigma_i}\left[p_t(\boldsymbol{\sigma})f_i(\boldsymbol{\sigma})\right] + T\frac{\partial^2}{\partial\sigma_i^2}p_t(\boldsymbol{\sigma}). \tag{2.131}$$

In the case of graded–response neurons, the continuous variable σ_i represents the membrane potential of neuron i, and (in their simplest form) the deterministic forces are given by

$$f_i(\boldsymbol{\sigma}) = J_{ij}\tanh[\gamma\sigma_j] - \sigma_i + \theta_i, \qquad \text{with} \qquad \gamma > 0,$$

and with the θ_i representing injected currents. Conventional notation is restored by putting $\sigma_i \to u_i$. Thus equation (2.128) specializes to

$$\dot{u}_i(t) = J_{ij} \tanh[\gamma u_j(t)] - u_i(t) + \theta_i + \eta_i(t). \tag{2.132}$$

One often chooses $T = 0$ (i.e., $\eta_i(t) = 0$), the rationale being that threshold noise is already assumed to have been incorporated via the nonlinearity in (2.132).

In our second example the variables σ_i represent the phases of coupled neural oscillators, with forces of the form

$$f_i(\boldsymbol{\sigma}) = J_{ij} \sin(\sigma_j - \sigma_i) + \omega_i.$$

Individual synapses J_{ij} now try to enforce either pair–wise synchronization ($J_{ij} > 0$), or pair–wise anti–synchronization ($J_{ij} < 0$), and the ω_i represent the natural frequencies of the individual oscillators. Conventional notation dictates $\sigma_i \to \xi_i$, giving [Coo01, SC00, SC01, CKS05]

$$\dot{\xi}_i(t) = \omega_i + J_{ij} \sin[\xi_j(t) - \xi_i(t)] + \eta_i(t). \tag{2.133}$$

2.3.2 Synaptic Symmetry and Lyapunov Functions

In the deterministic limit $T \to 0$ the rules (2.121) for networks of synchronously evolving binary neurons reduce to the deterministic map

$$\sigma_i(\ell + 1) = \operatorname{sgn}\left[h_i(\boldsymbol{\sigma}(\ell))\right]. \tag{2.134}$$

It turns out that for systems with symmetric interactions, $J_{ij} = J_{ji}$ for all (ij), one can construct a Lyapunov function, i.e., a function of $\boldsymbol{\sigma}$ which during the dynamics decreases monotonically and is bounded from below (see e.g., [Kha92]):

$$\text{Binary \& Parallel Dynamics:} \qquad L[\boldsymbol{\sigma}] = -\sum_i |h_i(\boldsymbol{\sigma})| - \sigma_i \theta_i. \tag{2.135}$$

Clearly, $L \geq -\sum_i [\sum_j |J_{ij}| + |\theta_i|] - \sum_i |\theta_i|$. During iteration of (2.134) we find:

$$\begin{aligned} L[\boldsymbol{\sigma}(\ell+1)] - L[\boldsymbol{\sigma}(\ell)] &= -\sum_i |h_i(\boldsymbol{\sigma}(\ell+1))| \\ &\quad + \sigma_i(\ell+1)[J_{ij}\sigma_j(\ell) + \theta_i] - \theta_i\left[\sigma_i(\ell+1) - \sigma_i(\ell)\right] \\ &= -\sum_i |h_i(\boldsymbol{\sigma}(\ell+1))| + \sigma_i(\ell) h_i(\boldsymbol{\sigma}(\ell+1)) \\ &= -\sum_i |h_i(\boldsymbol{\sigma}(\ell+1))| \left[1 - \sigma_i(\ell+2)\sigma_i(\ell)\right] \;\leq\; 0, \end{aligned}$$

where we have used (2.134) and $J_{ij} = J_{ji}$. So L decreases monotonically until a stage is reached where $\sigma_i(\ell + 2) = \sigma_i(\ell)$ for all i. Thus, with symmetric interactions this system will in the deterministic limit always end up in a

limit cycle with period ≤ 2. A similar result is found for networks with binary neurons and sequential dynamics. In the limit $T \to 0$ the rules (2.122) reduce to the map

$$\sigma_i(\ell+1) = \delta_{i,i_\ell}\mathrm{sgn}\left[h_i(\boldsymbol{\sigma}(\ell))\right] + [1-\delta_{i,i_\ell}]\sigma_i(\ell). \tag{2.136}$$

(in which we still have randomness in the choice of site to be updated). For systems with symmetric interactions and without self–interactions, i.e., $J_{ii} = 0$ for all i, we again find a Lyapunov function:

$$\text{Binary \& Sequential Dynamics:} \qquad L[\boldsymbol{\sigma}] = -\frac{1}{2}\sigma_i J_{ij}\sigma_j - \sigma_i\theta_i. \tag{2.137}$$

This quantity is bounded from below, $L \geq -\frac{1}{2}\sum_{ij}|J_{ij}| - \sum_i |\theta_i|$. Upon calling the site i_ℓ selected for update at step ℓ simply i, the change in L during iteration of (2.136) can be written as [Coo01, SC00, SC01, CKS05]:

$$\begin{aligned} L[\boldsymbol{\sigma}(\ell+1)] - L[\boldsymbol{\sigma}(\ell)] &= -\theta_i[\sigma_i(\ell+1) - \sigma_i(\ell)] \\ &\quad - \frac{1}{2}J_{ik}[\sigma_i(\ell+1)\sigma_k(\ell+1) - \sigma_i(\ell)\sigma_k(\ell)] \\ &\quad - \frac{1}{2}J_{ji}[\sigma_j(\ell+1)\sigma_i(\ell+1) - \sigma_j(\ell)\sigma_i(\ell)] \\ &= [\sigma_i(\ell) - \sigma_i(\ell+1)][J_{ij}\sigma_j(\ell) + \theta_i] \\ &= -|h_i(\boldsymbol{\sigma}(\ell))|\,[1 - \sigma_i(\ell)\sigma_i(\ell+1)] \;\leq\; 0. \end{aligned}$$

Here we used (2.136), $J_{ij} = J_{ji}$, and absence of self–interactions. Thus L decreases monotonically until $\sigma_i(t+1) = \sigma_i(t)$ for all i. With symmetric synapses, but without diagonal terms, the sequentially evolving binary neurons system will in the deterministic limit always end up in a stationary state.

Now, one can derive similar results for models with continuous variables. Firstly, in the deterministic limit the graded–response equations (2.132) simplify to

$$\dot{u}_i(t) = J_{ij}\tanh[\gamma u_j(t)] - u_i(t) + \theta_i. \tag{2.138}$$

Symmetric networks again admit a Lyapunov function (without a need to eliminate self-interactions):

$$\text{Graded–Response Dynamics :} \qquad L[\mathbf{u}] = -\frac{1}{2}J_{ij}\tanh(\gamma u_i)\tanh(\gamma u_j) + \sum_i \left[\gamma\int_0^{u_i} dv\; v[1-\tanh^2(\gamma v)] - \theta_i\tanh(\gamma u_i)\right].$$

Clearly, $L \geq -\frac{1}{2}\sum_{ij}|J_{ij}| - \sum_i|\theta_i|$; the term in $L[\mathbf{u}]$ with the integral is non–negative. During the noise–free dynamics (2.138) one can use the identity

$$\frac{\partial L}{\partial u_i} = -\gamma\dot{u}_i[1-\tanh^2(\gamma u_i)],$$

valid only when $J_{ij} = J_{ji}$, to derive [Coo01, SC00, SC01, CKS05]

$$\dot{L} = \frac{\partial L}{\partial u_i}\dot{u}_i = -\gamma \sum_i [1 - \tanh^2(\gamma u_i)]\ \dot{u}_i^2 \leq 0.$$

Again L is found to decrease monotonically, until $\dot{u}_i = 0$ for all i, i.e., until we are at a fixed–point.

The coupled oscillator equations (2.133) reduce in the *noise–free limit* to

$$\dot{\xi}_i(t) = \omega_i + J_{ij} \sin[\xi_j(t) - \xi_i(t)]. \tag{2.139}$$

Note that self-interactions J_{ii} always drop out automatically. For symmetric oscillator networks, a construction of the type followed for the graded–response equations would lead us to propose

$$\text{Coupled Oscillators Dynamics:} \qquad L[\boldsymbol{\xi}] = -\frac{1}{2} J_{ij} \cos[\xi_i - \xi_j] - \omega_i \xi_i. \tag{2.140}$$

This function decreases monotonically, due to $\partial L/\partial \xi_i = -\dot{\xi}_i$:

$$\dot{L} = \frac{\partial L}{\partial \xi_i}\dot{\xi}_i = -\sum_i \dot{\xi}_i^2 \leq 0.$$

Actually, (2.139) describes gradient descent on the surface $L[\boldsymbol{\xi}]$. However, due to the term with the natural frequencies ω_i the function $L[\boldsymbol{\xi}]$ is not bounded, so it cannot be a Lyapunov function. This could have been expected; when $J_{ij} = 0$ for all (i, j), for instance, one finds continually increasing phases, $\xi_i(t) = \xi_i(0) + \omega_i t$. Removing the ω_i, in contrast, gives the bound $L \geq -\sum_j |J_{ij}|$. Now the system must go to a fixed-point. In the special case $\omega_i = \omega$ (N identical natural frequencies) we can transform away the ω_i by putting $\xi(t) = \tilde{\xi}_i(t) + \omega t$, and find the relative phases $\tilde{\xi}_i$ to go to a fixed-point.

2.3.3 Detailed Balance and Equilibrium Statistical Mechanics

The results got above indicate that networks with symmetric synapses are a special class [Coo01, SC00, SC01, CKS05]. We now show how synaptic symmetry is closely related to the detailed balance property, and derive a number of consequences. An ergodic Markov chain of the form (2.124,2.125), i.e.,

$$p_{\ell+1}(\boldsymbol{\sigma}) = \sum_{\boldsymbol{\sigma}'} W\left[\boldsymbol{\sigma}; \boldsymbol{\sigma}'\right] p_\ell(\boldsymbol{\sigma}'), \tag{2.141}$$

is said to obey detailed balance if its (unique) stationary solution $p_\infty(\boldsymbol{\sigma})$ has the property

$$W\left[\boldsymbol{\sigma}; \boldsymbol{\sigma}'\right] p_\infty(\boldsymbol{\sigma}') = W\left[\boldsymbol{\sigma}'; \boldsymbol{\sigma}\right] p_\infty(\boldsymbol{\sigma}), \qquad (\text{for all } \boldsymbol{\sigma}, \boldsymbol{\sigma}'). \tag{2.142}$$

All $p_\infty(\boldsymbol{\sigma})$ which satisfy (2.142) are stationary solutions of (2.141), this is easily verified by substitution. The converse is not true. Detailed balance states that, in addition to $p_\infty(\boldsymbol{\sigma})$ being stationary, one has *equilibrium*: there is no net probability current between any two microscopic system states.

It is not a trivial matter to investigate systematically for which choices of the threshold noise distribution $w(\eta)$ and the synaptic matrix $\{J_{ij}\}$ detailed balance holds. It can be shown that, apart from trivial cases (e.g., systems with self–interactions only) a Gaussian distribution $w(\eta)$ will not support detailed balance. Here we will work out details only for the choice $w(\eta) = \frac{1}{2}[1 - \tanh^2(\eta)]$, and for $T > 0$ (where both discrete systems are ergodic). For parallel dynamics the transition matrix is given in (2.124), now with $g[z] = \tanh[z]$, and the detailed balance condition (2.142) becomes

$$\frac{\mathrm{e}^{\beta\sigma_i h_i(\boldsymbol{\sigma}')} p_\infty(\boldsymbol{\sigma}')}{\prod_i \cosh[\beta h_i(\boldsymbol{\sigma}')]} = \frac{\mathrm{e}^{\beta\sigma_i' h_i(\boldsymbol{\sigma})} p_\infty(\boldsymbol{\sigma})}{\prod_i \cosh[\beta h_i(\boldsymbol{\sigma})]}, \qquad \text{(for all } \boldsymbol{\sigma}, \boldsymbol{\sigma}'\text{)}. \tag{2.143}$$

All $p_\infty(\boldsymbol{\sigma})$ are non-zero (ergodicity), so we may safely put

$$p_\infty(\boldsymbol{\sigma}) = \mathrm{e}^{\beta[\theta_i\sigma_i + K(\boldsymbol{\sigma})]} \prod_i \cosh[\beta h_i(\boldsymbol{\sigma})],$$

which, in combination with definition (2.121), simplifies the detailed balance condition to:

$$K(\boldsymbol{\sigma}) - K(\boldsymbol{\sigma}') = \sigma_i \left[J_{ij} - J_{ji}\right] \sigma_j', \qquad \text{(for all } \boldsymbol{\sigma}, \boldsymbol{\sigma}'\text{)}. \tag{2.144}$$

Averaging (2.144) over all possible $\boldsymbol{\sigma}'$ gives $K(\boldsymbol{\sigma}) = \langle K(\boldsymbol{\sigma}')\rangle_{\boldsymbol{\sigma}'}$ for all $\boldsymbol{\sigma}$, i.e., K is a constant, whose value follows from normalizing $p_\infty(\boldsymbol{\sigma})$. So, if detailed balance holds the equilibrium distribution must be [Coo01, SC00, SC01, CKS05]:

$$p_{\mathrm{eq}}(\boldsymbol{\sigma}) \;\sim\; \mathrm{e}^{\beta\theta_i\sigma_i} \prod_i \cosh[\beta h_i(\boldsymbol{\sigma})]. \tag{2.145}$$

For symmetric systems detailed balance indeed holds: (2.145) solves (2.143), since $K(\boldsymbol{\sigma}) = K$ solves the reduced problem (2.144). For non-symmetric systems, however, there can be no equilibrium. For $K(\boldsymbol{\sigma}) = K$ the condition (2.144) becomes $\sum_{ij} \sigma_i \left[J_{ij} - J_{ji}\right] \sigma_j' = 0$ for all $\boldsymbol{\sigma}, \boldsymbol{\sigma}' \in \{-1, 1\}^N$. For $N \geq 2$ the vector pairs $(\boldsymbol{\sigma}, \boldsymbol{\sigma}')$ span the space of all $N \times N$ matrices, so $J_{ij} - J_{ji}$ must be zero. For $N = 1$ there simply exists no non-symmetric synaptic matrix. In conclusion: for binary networks with parallel dynamics, interaction symmetry implies detailed balance, and vice versa.

For sequential dynamics, with $w(\eta) = \frac{1}{2}[1 - \tanh^2(\eta)]$, the transition matrix is given by (2.125) and the detailed balance condition (2.142) simplifies to

$$\frac{\mathrm{e}^{\beta\sigma_i h_i(F_i\boldsymbol{\sigma})} p_\infty(F_i\boldsymbol{\sigma})}{\cosh\left[\beta h_i(F_i\boldsymbol{\sigma})\right]} = \frac{\mathrm{e}^{-\beta\sigma_i h_i(\boldsymbol{\sigma})} p_\infty(\boldsymbol{\sigma})}{\cosh\left[\beta h_i(\boldsymbol{\sigma})\right]}, \qquad \text{(for all } \boldsymbol{\sigma}, i\text{)}.$$

Self–interactions J_{ii}, inducing $h_i(F_i\boldsymbol{\sigma}) \neq h_i(\boldsymbol{\sigma})$, complicate matters. Therefore we first consider systems where all $J_{ii} = 0$. All stationary probabilities $p_\infty(\boldsymbol{\sigma})$ being non–zero (ergodicity), we may write:

$$p_\infty(\boldsymbol{\sigma}) = \mathrm{e}^{\beta[\theta_i\sigma_i + \frac{1}{2}\sigma_i J_{ij}\sigma_j + K(\boldsymbol{\sigma})]}. \tag{2.146}$$

Using relations like

$$J_{kl}F_i(\sigma_k\sigma_l) = J_{kl}\sigma_k\sigma_l - 2\sigma_i\left[J_{ik} + J_{ki}\right]\sigma_k,$$

we can simplify the detailed balance condition to

$$K(F_i\boldsymbol{\sigma}) - K(\boldsymbol{\sigma}) = \sigma_i\left[J_{ik} - J_{ki}\right]\sigma_k, \qquad (\text{for all } \boldsymbol{\sigma}, i).$$

If to this expression we apply the general identity

$$[1 - F_i]\, f(\boldsymbol{\sigma}) = 2\sigma_i\langle\sigma_i f(\boldsymbol{\sigma})\rangle_{\sigma_i},$$

we find for $i \neq j$ [Coo01, SC00, SC01, CKS05]:

$$K(\boldsymbol{\sigma}) = -2\sigma_i\sigma_j\left[J_{ij} - J_{ji}\right], \qquad (\text{for all } \boldsymbol{\sigma} \text{ and all } i \neq j).$$

The left–hand side is symmetric under permutation of the pair (i, j), which implies that the interaction matrix must also be symmetric: $J_{ij} = J_{ji}$ for all (i, j). We now find the trivial solution $K(\boldsymbol{\sigma}) = K$ (constant), detailed balance holds and the corresponding equilibrium distribution is

$$p_{\mathrm{eq}}(\boldsymbol{\sigma}) \;\sim\; \mathrm{e}^{-\beta H(\boldsymbol{\sigma})}, \qquad H(\boldsymbol{\sigma}) = -\frac{1}{2}\sigma_i J_{ij}\sigma_j - \theta_i\sigma_i.$$

In conclusion: for binary networks with sequential dynamics, but without self–interactions, interaction symmetry implies detailed balance, and vice versa. In the case of self–interactions the situation is more complicated. However, here one can still show that non-symmetric models with detailed balance must be pathological, since the requirements can be met only for very specific choices for the $\{J_{ij}\}$.

Now, let us turn to the question of when we find microscopic equilibrium (stationarity without probability currents) in continuous models described by a Fokker–Planck equation (2.131). Note that (2.131) can be seen as a continuity equation for the density of a conserved quantity:

$$\dot{p}_t(\boldsymbol{\sigma}) + \frac{\partial}{\partial\sigma_i}J_i(\boldsymbol{\sigma}, t) = 0.$$

The components $J_i(\boldsymbol{\sigma}, t)$ of the current density are given by

$$J_i(\boldsymbol{\sigma}, t) = [f_i(\boldsymbol{\sigma}) - T\frac{\partial}{\partial\sigma_i}]p_t(\boldsymbol{\sigma}).$$

Stationary distributions $p_\infty(\boldsymbol{\sigma})$ are those which give $\sum_i \frac{\partial}{\partial\sigma_i} J_i(\boldsymbol{\sigma},\infty) = 0$ (divergence-free currents). Detailed balance implies the stronger statement $J_i(\boldsymbol{\sigma},\infty) = 0$ for all i (zero currents), so

$$f_i(\boldsymbol{\sigma}) = T\frac{\partial \log p_\infty(\boldsymbol{\sigma})}{\partial\sigma_i}, \qquad \text{or}$$

$$f_i(\boldsymbol{\sigma}) = -\frac{\partial H(\boldsymbol{\sigma})}{\partial\sigma_i}, \qquad p_\infty(\boldsymbol{\sigma}) \;\sim\; \mathrm{e}^{-\beta H(\boldsymbol{\sigma})}, \tag{2.147}$$

for some $H(\boldsymbol{\sigma})$, i.e., the forces $f_i(\boldsymbol{\sigma})$ must be conservative. However, one can have conservative forces without a normalizable equilibrium distribution. Just take $H(\boldsymbol{\sigma}) = 0$, i.e., $f_i(\boldsymbol{\sigma},t) = 0$: here we have $p_{\mathrm{eq}}(\boldsymbol{\sigma}) = C$, which is not normalizable for $\boldsymbol{\sigma} \in \mathbb{R}^N$. For this particular case equation (2.131) is solved easily:

$$p_t(\boldsymbol{\sigma}) = [4\pi T t]^{-N/2} \int d\boldsymbol{\sigma}'\; p_0(\boldsymbol{\sigma}')\mathrm{e}^{-[\boldsymbol{\sigma}-\boldsymbol{\sigma}']^2/4Tt},$$

so the limit $\lim_{t\to\infty} p_t(\boldsymbol{\sigma})$ does not exist. One can prove the following (see e.g., [Zin93]). If the forces are conservative and if $p_\infty(\boldsymbol{\sigma}) \sim \mathrm{e}^{-\beta H(\boldsymbol{\sigma})}$ is normalizable, then it is the unique stationary solution of the Fokker–Planck equation, to which the system converges for all initial distributions $p_0 \in L^1[\mathbb{R}^N]$ which obey $\int_{\mathbb{R}^N} d\boldsymbol{\sigma}\; \mathrm{e}^{\beta H(\boldsymbol{\sigma})} p_0^2(\boldsymbol{\sigma}) < \infty$.

Note that conservative forces must obey [Coo01, SC00, SC01, CKS05]

$$\frac{\partial f_i(\boldsymbol{\sigma})}{\partial\sigma_j} - \frac{\partial f_j(\boldsymbol{\sigma})}{\partial\sigma_i} = 0, \qquad (\text{for all } \boldsymbol{\sigma} \text{ and all } i \neq j). \tag{2.148}$$

In the graded–response equations (2.138) the deterministic forces are

$$f_i(\mathbf{u}) = J_{ij}\tanh[\gamma u_j] - u_i + \theta_i, \qquad \text{where}$$
$$\frac{\partial f_i(\mathbf{u})}{\partial u_j} - \frac{\partial f_j(\mathbf{u})}{\partial u_i} = \gamma\{J_{ij}[1-\tanh^2[\gamma u_j] - J_{ji}[1-\tanh^2[\gamma u_i]\}.$$

At $\mathbf{u} = \mathbf{0}$ this reduces to $J_{ij} - J_{ji}$, i.e., the interaction matrix must be symmetric. For symmetric matrices we find away from $\mathbf{u} = \mathbf{0}$:

$$\frac{\partial f_i(\mathbf{u})}{\partial u_j} - \frac{\partial f_j(\mathbf{u})}{\partial u_i} = \gamma J_{ij}\{\tanh^2[\gamma u_i] - \tanh^2[\gamma u_j]\}.$$

The only way for this to be zero for any $\mathbf{u}$ is by having $J_{ij} = 0$ for all $i \neq j$, i.e., all neurons are disconnected (in this trivial case the system (2.138) does indeed obey detailed balance). Network models of interacting graded–response neurons of the type (2.138) apparently never reach equilibrium, they will always violate detailed balance and exhibit microscopic probability currents. In the case of coupled oscillators (2.133), where the deterministic forces are

$$f_i(\boldsymbol{\xi}) = J_{ij}\sin[\xi_j - \xi_i] + \omega_i,$$

one finds the left-hand side of condition (2.148) to give

$$\frac{\partial f_i(\boldsymbol{\xi})}{\partial \xi_j} - \frac{\partial f_j(\boldsymbol{\xi})}{\partial \xi_i} = [J_{ij} - J_{ji}] \cos[\xi_j - \xi_i].$$

Requiring this to be zero for any $\boldsymbol{\xi}$ gives the condition $J_{ij} = J_{ji}$ for any $i \neq j$. We have already seen that symmetric oscillator networks indeed have conservative forces:

$$f_i(\boldsymbol{\xi}) = -\partial H(\boldsymbol{\xi})/\partial \xi_i, \qquad \text{with} \qquad H(\boldsymbol{\xi}) = -\frac{1}{2} J_{ij} \cos[\xi_i - \xi_j] - \omega_i \xi_i.$$

If in addition we choose all $\omega_i = 0$ the function $H(\boldsymbol{\sigma})$ will also be bounded from below, and, although $p_\infty(\boldsymbol{\xi}) \sim \mathrm{e}^{-\beta H(\boldsymbol{\xi})}$ is still not normalizable on $\boldsymbol{\xi} \in \mathbb{R}^N$, the full $2\pi-$periodicity of the function $H(\boldsymbol{\sigma})$ now allows us to identify $\xi_i + 2\pi \equiv \xi_i$ for all i, so that now $\boldsymbol{\xi} \in [-\pi, \pi]^N$ and $\int d\boldsymbol{\xi}\ \mathrm{e}^{-\beta H(\boldsymbol{\xi})}$ does exist. Thus symmetric coupled oscillator networks with zero natural frequencies obey detailed balance. In the case of non-zero natural frequencies, in contrast, detailed balance does not hold.

The above results establish the link with *equilibrium statistical mechanics* (see e.g., [Yeo92, PB94]). For binary systems with symmetric synapses (in the sequential case: without self-interactions) and with threshold noise distributions of the form

$$w(\eta) = \frac{1}{2}[1 - \tanh^2(\eta)],$$

detailed balance holds and we know the equilibrium distributions. For sequential dynamics it has the Boltzmann form (2.147) and we can apply standard equilibrium statistical mechanics. The parameter β can formally be identified with the inverse 'temperature' in equilibrium, $\beta = T^{-1}$, and the function $H(\boldsymbol{\sigma})$ is the usual *Ising–spin Hamiltonian*. In particular we can define the *partition function* Z and the *free energy* F [Coo01, SC00, SC01, CKS05]:

$$p_{\mathrm{eq}}(\boldsymbol{\sigma}) = \frac{1}{Z} \mathrm{e}^{-\beta H(\boldsymbol{\sigma})}, \qquad H(\boldsymbol{\sigma}) = -\frac{1}{2} \sigma_i J_{ij} \sigma_j - \theta_i \sigma_i, \tag{2.149}$$

$$Z = \sum_{\boldsymbol{\sigma}} \mathrm{e}^{-\beta H(\boldsymbol{\sigma})}, \qquad F = -\beta^{-1} \log Z. \tag{2.150}$$

The free energy can be used as the partition function for equilibrium averages. Taking derivatives with respect to external fields θ_i and interactions J_{ij}, for instance, produces $\langle \sigma_i \rangle = -\partial F/\partial \theta_i$ and $\langle \sigma_i \sigma_j \rangle = -\partial F/\partial J_{ij}$, whereas equilibrium averages of arbitrary state variable $f(\boldsymbol{\sigma})$ can be obtained by adding suitable partition terms to the Hamiltonian:

$$H(\boldsymbol{\sigma}) \to H(\boldsymbol{\sigma}) + \lambda f(\boldsymbol{\sigma}), \qquad \langle f \rangle = \lim_{\lambda \to 0} \frac{\partial F}{\partial \lambda}.$$

In the parallel case (2.145) we can again formally write the equilibrium probability distribution in the Boltzmann form [Per84] and define a corresponding partition function $\tilde{Z}$ and a free energy $\tilde{F}$:

$$p_{\rm eq}(\boldsymbol{\sigma}) = \frac{1}{Z}{\rm e}^{-\beta\tilde{H}(\boldsymbol{\sigma})}, \qquad \tilde{H}(\boldsymbol{\sigma}) = -\theta_i\sigma_i - \frac{1}{\beta}\sum_i \log 2\cosh[\beta h_i(\boldsymbol{\sigma})], \quad (2.151)$$

$$\tilde{Z} = \sum_{\boldsymbol{\sigma}} {\rm e}^{-\beta\tilde{H}(\boldsymbol{\sigma})}, \qquad \tilde{F} = -\beta^{-1}\log\tilde{Z}, \quad (2.152)$$

which again serve to generate averages: $\tilde{H}(\boldsymbol{\sigma}) \to \tilde{H}(\boldsymbol{\sigma}) + \lambda f(\boldsymbol{\sigma})$, $\langle f\rangle = \lim_{\lambda\to 0}\partial\tilde{F}/\partial\lambda$. However, standard thermodynamic relations involving derivation with respect to β need no longer be valid, and derivation with respect to fields or interactions generates different types of averages, such as [Coo01, SC00, SC01, CKS05]

$$-\frac{\partial\tilde{F}}{\partial\theta_i} = \langle\sigma_i\rangle + \langle\tanh[\beta h_i(\boldsymbol{\sigma})]\rangle, \qquad -\frac{\partial\tilde{F}}{\partial J_{ii}} = \langle\sigma_i\tanh[\beta h_i(\boldsymbol{\sigma})]\rangle,$$

$$i\neq j: \qquad \frac{\partial\tilde{F}}{\partial J_{ii}} = \langle\sigma_i\tanh[\beta h_j(\boldsymbol{\sigma})]\rangle + \langle\sigma_j\tanh[\beta h_i(\boldsymbol{\sigma})]\rangle.$$

One can use $\langle\sigma_i\rangle = \langle\tanh[\beta h_i(\boldsymbol{\sigma})]\rangle$, which can be derived directly from the equilibrium equation $p_{\rm eq}(\boldsymbol{\sigma}) = \sum_{\boldsymbol{\sigma}'} W[\boldsymbol{\sigma};\boldsymbol{\sigma}']p_{\rm eq}(\boldsymbol{\sigma})$, to simplify the first of these identities.

A connected network of graded–response neurons can never be in an equilibrium state, so our only model example with continuous neuronal variables for which we can set up the equilibrium statistical mechanics formalism is the system of coupled oscillators (2.133) with symmetric synapses and absent (or uniform) natural frequencies ω_i. If we define the phases as $\xi_i \in [-\pi,\pi]$ we have again an equilibrium distribution of the Boltzmann form, and we can define the standard thermodynamic quantities:

$$p_{\rm eq}(\boldsymbol{\xi}) = \frac{1}{Z}{\rm e}^{-\beta H(\boldsymbol{\xi})}, \qquad H(\boldsymbol{\xi}) = -\frac{1}{2}J_{ij}\cos[\xi_i - \xi_j], \quad (2.153)$$

$$Z = \int_{-\pi}^{\pi}\cdots\int_{-\pi}^{\pi} d\boldsymbol{\xi}\ {\rm e}^{-\beta H(\boldsymbol{\xi})}, \qquad F = -\beta^{-1}\log Z. \quad (2.154)$$

These generate equilibrium averages in the usual manner. For instance

$$\langle\cos[\xi_i - \xi_j]\rangle = -\frac{\partial F}{\partial J_{ij}},$$

whereas averages of arbitrary state variables $f(\boldsymbol{\xi})$ follow, as before, upon introducing suitable partition terms:

$$H(\boldsymbol{\xi}) \to H(\boldsymbol{\xi}) + \lambda f(\boldsymbol{\xi}), \qquad \langle f\rangle = \lim_{\lambda\to 0}\frac{\partial F}{\partial\lambda}.$$

2.3.4 Simple Recurrent Networks with Binary Neurons

Networks with Uniform Synapses

We now turn to a simple toy model to show how equilibrium statistical mechanics is used for solving neural network models, and to illustrate similarities and differences between the different dynamics types [Coo01, SC00, SC01, CKS05]. We choose uniform infinite-range synapses and zero external fields, and calculate the free energy for the binary systems (2.121,2.122), parallel and sequential, and with threshold–noise distribution $w(\eta) = \frac{1}{2}[1 - \tanh^2(\eta)]$:

$$J_{ij} = J_{ji} = J/N, \qquad (i \neq j), \qquad J_{ii} = \theta_i = 0, \qquad (\text{for all } i).$$

The free energy is an extensive object, $\lim_{N\to\infty} F/N$ is finite. For the models (2.121,2.122) we now get:

Binary & Sequential Dynamics:

$$\lim_{N\to\infty} F/N = -\lim_{N\to\infty} (\beta N)^{-1} \log \sum_{\boldsymbol{\sigma}} \mathrm{e}^{\beta N\left[\frac{1}{2}Jm^2(\boldsymbol{\sigma})\right]},$$

Binary & Parallel Dynamics:

$$\lim_{N\to\infty} \tilde{F}/N = -\lim_{N\to\infty} (\beta N)^{-1} \log \sum_{\boldsymbol{\sigma}} \mathrm{e}^{N[\log 2 \cosh[\beta J m(\boldsymbol{\sigma})]]},$$

with the average activity $m(\boldsymbol{\sigma}) = \frac{1}{N}\sum_k \sigma_k$. We have to count the number of states $\boldsymbol{\sigma}$ with a prescribed average activity $m = 2n/N - 1$ (n is the number of neurons i with $\sigma_i = 1$), in expressions of the form

$$\begin{aligned}
\frac{1}{N} \log \sum_{\boldsymbol{\sigma}} \mathrm{e}^{NU[m(\boldsymbol{\sigma})]} &= \frac{1}{N} \log \sum_{n=0}^{N} \binom{N}{n} \mathrm{e}^{NU[2n/N-1]} \\
&= \frac{1}{N} \log \int_{-1}^{1} dm \; \mathrm{e}^{N[\log 2 - c^*(m) + U[m]]}, \\
\lim_{N\to\infty} \frac{1}{N} \log \sum_{\boldsymbol{\sigma}} \mathrm{e}^{NU[m(\boldsymbol{\sigma})]} &= \log 2 + \max_{m\in[-1,1]} \{U[m] - c^*(m)\},
\end{aligned}$$

with the *entropic function*

$$c^*(m) = \frac{1}{2}(1+m)\log(1+m) + \frac{1}{2}(1-m)\log(1-m).$$

In order to get there we used Stirling's formula to get the leading term of the factorials (only terms which are exponential in N survive the limit $N \to \infty$), we converted (for $N \to \infty$) the summation over n into an integration over $m = 2n/N - 1 \in [-1, 1]$, and we carried out the integral over m via *saddle–point integration* (see e.g., [Per92]). This leads to a saddle–point problem whose solution gives the free energies [Coo01, SC00, SC01, CKS05]:

$$\lim_{N\to\infty} F/N = \min_{m\in[-1,1]} f_{\rm seq}(m), \qquad \beta f_{\rm seq}(m) = c^*(m) - \log 2 - \frac{1}{2}\beta J m^2. \tag{2.155}$$

$$\lim_{N\to\infty} \tilde{F}/N = \min_{m\in[-1,1]} f_{\rm par}(m), \qquad \beta f_{\rm par}(m) = c^*(m) - 2\log 2 - \log\cosh[\beta J m]. \tag{2.156}$$

The equations from which to solve the minima are easily got by differentiation, using $\frac{d}{dm}c^*(m) = \tanh^{-1}(m)$. For sequential dynamics we find

$$\text{Binary \& Sequential Dynamics:} \qquad m = \tanh[\beta J m], \tag{2.157}$$

which is the so–called *Curie–Weiss law*. For parallel dynamics we find

$$m = \tanh\left[\beta J \tanh[\beta J m]\right].$$

One finds that the solutions of the latter equation again obey a Curie–Weiss law. The definition $\hat{m} = \tanh[\beta|J|m]$ transforms it into the coupled equations $m = \tanh[\beta|J|\hat{m}]$ and $\hat{m} = \tanh[\beta|J|m]$, from which we derive

$$0 \leq [m-\hat{m}]^2 = [m-\hat{m}]\left[\tanh[\beta|J|\hat{m}] - \tanh[\beta|J|m]\right] \leq 0.$$

Since $\tanh[\beta|J|m]$ is a monotonically increasing function of m, this implies $\hat{m} = m$, so

$$\text{Binary \& Parallel Dynamics:} \qquad m = \tanh[\beta|J|m]. \tag{2.158}$$

Our study of the toy models has thus been reduced to analyzing the nonlinear equations (2.157) and (2.158). If $J \geq 0$ (excitation) the two types of dynamics lead to the same behavior. At high noise levels, $T > J$, both minimisation problems are solved by $m = 0$, describing a disorganized (paramagnetic) state. This can be seen upon writing the right–hand side of (2.157) in integral form [Coo01, SC00, SC01, CKS05]:

$$m^2 = m\tanh[\beta J m] = \beta J m^2 \int_0^1 dz\ [1 - \tanh^2[\beta J m z]] \leq \beta J m^2.$$

So $m^2[1-\beta J] \leq 0$, which gives $m = 0$ as soon as $\beta J < 1$. A phase transition occurs at $T = J$ (a bifurcation of non–trivial solutions of (2.157)), and for $T < J$ the equations for m are solved by the two non-zero solutions of (2.157), describing a state where either all neurons tend to be firing ($m > 0$) or where they tend to be quiet ($m < 0$). This becomes clear when we expand (2.157) for small m: $m = \beta J m + \mathcal{O}(m^3)$, so precisely at $\beta J = 1$ one finds a de–stabilization of the trivial solution $m = 0$, together with the creation of (two) stable non–trivial ones. Furthermore, using the identity $c^*(\tanh x) = x\tanh x - \log\cosh x$, we get from (2.155,2.156) the relation $\lim_{N\to\infty}\tilde{F}/N = 2\lim_{N\to\infty} F/N$. For $J < 0$ (inhibition), however, the two types of dynamics give quite different results. For sequential dynamics the relevant minimum is located at $m = 0$ (the paramagnetic state). For parallel dynamics, the minimization problem is

invariant under $J \to -J$, so the behavior is again of the Curie-Weiss type, with a paramagnetic state for $T > |J|$, a phase transition at $T = |J|$, and order for $T < |J|$. This difference between the two types of dynamics for $J < 0$ is explained by studying dynamics. For the present (toy) model in the limit $N \to \infty$ the average activity evolves in time according to the deterministic laws [Coo01, SC00, SC01, CKS05]

$$\dot{m} = \tanh[\beta J m] - m, \qquad m(t+1) = \tanh[\beta J m(t)],$$

for sequential and parallel dynamics, respectively. For $J < 0$ the sequential system always decays towards the trivial state $m = 0$, whereas for sufficiently large β the parallel system enters the stable limit–cycle $m(t) = M_\beta(-1)^t$, where M_β is the non-zero solution of (2.158). The concepts of 'distance' and 'local minima' are quite different for the two dynamics types; in contrast to the sequential case, parallel dynamics allows the system to make the transition $m \to -m$ in equilibrium.

Phenomenology of Hopfield Models

Recall that the Hopfield model [Hop82] represents a network of binary neurons of the type (2.121,2.122), with threshold noise $w(\eta) = \frac{1}{2}[1 - \tanh^2(\eta)]$, and with a specific recipe for the synapses J_{ij} aimed at storing patterns, motivated by suggestions made in the late nineteen-forties [Heb49]. The original model was in fact defined more narrowly, as the zero noise limit of the system (2.122), but the term has since then been accepted to cover a larger network class. Let us first consider the simplest case and try to store a single pattern $\boldsymbol{\xi} \in \{-1, 1\}^N$ in noise–less infinite–range binary networks. Appealing candidates for interactions and thresholds would be $J_{ij} = \xi_i \xi_j$ and $\theta_i = 0$ (for sequential dynamics we put $J_{ii} = 0$ for all i). With this choice the Lyapunov function (2.137) becomes:

$$L_{\rm seq}[\boldsymbol{\sigma}] = \frac{1}{2}N - \frac{1}{2}[\xi_i \sigma_i]^2.$$

This system indeed reconstructs dynamically the original pattern $\boldsymbol{\xi}$ from an input vector $\boldsymbol{\sigma}(0)$, at least for sequential dynamics. However, *en passant* we have created an additional attractor: the state -$\boldsymbol{\xi}$. This property is shared by all binary models in which the external fields are zero, where the Hamiltonians $H(\boldsymbol{\sigma})$ (2.149) and $\tilde{H}(\boldsymbol{\sigma})$ (2.151) are invariant under an overall sign change $\boldsymbol{\sigma} \to -\boldsymbol{\sigma}$. A second feature common to several (but not all) attractor neural networks is that *each* initial state will lead to pattern reconstruction, even nonsensical (random) ones.

The Hopfield model is got by generalizing the previous simple one-pattern recipe to the case of an arbitrary number p of binary patterns $\boldsymbol{\xi}^\mu = (\xi_1^\mu, \ldots, \xi_N^\mu) \in \{-1, 1\}^N$ [Coo01, SC00, SC01, CKS05]:

$$J_{ij} = \frac{1}{N}\xi_i^\mu \xi_j^\mu, \qquad \theta_i = 0 \qquad (\text{for all } i; \;\; \mu = 1, ..., p), \tag{2.159}$$
$$(\text{sequential dynamics}: \;\; J_{ii} \to 0, \quad \text{for all } i).$$

The prefactor N^{-1} has been inserted to ensure that the limit $N \to \infty$ will exist in future expressions. The process of interest is that where, triggered by correlation between the initial state and a stored pattern $\boldsymbol{\xi}^\lambda$, the state vector $\boldsymbol{\sigma}$ evolves towards $\boldsymbol{\xi}^\lambda$. If this happens, pattern $\boldsymbol{\xi}^\lambda$ is said to be recalled. The similarity between a state vector and the stored patterns is measured by so–called *Hopfield overlaps*

$$m_\mu(\boldsymbol{\sigma}) = \frac{1}{N}\xi_i^\mu \sigma_i. \tag{2.160}$$

The Hopfield model represents as an associative memory, in which the recall process is described in terms of overlaps.

Analysis of Hopfield Models Away From Saturation

A binary Hopfield network with parameters given by (2.159) obeys detailed balance, and the Hamiltonian $H(\boldsymbol{\sigma})$ (2.149) (corresponding to sequential dynamics) and the pseudo-Hamiltonian $\tilde{H}(\boldsymbol{\sigma})$ (2.151) (corresponding to parallel dynamics) become [Coo01, SC00, SC01, CKS05]

$$\begin{aligned} H(\boldsymbol{\sigma}) &= -\frac{1}{2}N\sum_{\mu=1}^{p} m_\mu^2(\boldsymbol{\sigma}) + \frac{1}{2}p, \\ \tilde{H}(\boldsymbol{\sigma}) &= -\frac{1}{\beta}\sum_i \log 2\cosh[\beta\xi_i^\mu m_\mu(\boldsymbol{\sigma})], \end{aligned} \tag{2.161}$$

with the overlaps (2.160). Solving the statics implies calculating the free energies F and $\tilde{F}$:

$$F = -\frac{1}{\beta}\log\sum_{\boldsymbol{\sigma}} \mathrm{e}^{-\beta H(\boldsymbol{\sigma})}, \qquad \tilde{F} = -\frac{1}{\beta}\log\sum_{\boldsymbol{\sigma}} \mathrm{e}^{-\beta \tilde{H}(\boldsymbol{\sigma})}.$$

Upon introducing the short-hand notation $\mathbf{m} = (m_1, \ldots, m_p)$ and $\boldsymbol{\xi}_i = (\xi_i^1, \ldots, \xi_i^p)$, both free energies can be expressed in terms of the density of states $\mathcal{D}(\mathbf{m}) = 2^{-N}\sum_{\boldsymbol{\sigma}} \delta[\mathbf{m} - \mathbf{m}(\boldsymbol{\sigma})]$:

$$F/N = -\frac{1}{\beta}\log 2 - \frac{1}{\beta N}\log\int d\mathbf{m}\ \mathcal{D}(\mathbf{m})\ \mathrm{e}^{\frac{1}{2}\beta N\mathbf{m}^2} + \frac{p}{2N}, \tag{2.162}$$

$$\tilde{F}/N = -\frac{1}{\beta}\log 2 - \frac{1}{\beta N}\log\int d\mathbf{m}\ \mathcal{D}(\mathbf{m})\ \mathrm{e}^{\sum_{i=1}^N \log 2\cosh[\beta\boldsymbol{\xi}_i\cdot\mathbf{m}]}, \tag{2.163}$$

using $\int d\mathbf{m}\ \delta[\mathbf{m} - \mathbf{m}(\boldsymbol{\sigma})] = 1$. In order to proceed, we need to specify how the number of patterns p scales with the system size N. In this section we will follow [AGS85] (equilibrium analysis following sequential dynamics) and [FK88] (equilibrium analysis following parallel dynamics), and assume p to be finite. One can now easily calculate the leading contribution to the density

of states, using the integral representation of the $\delta-$function and keeping in mind that according to (2.162,2.163) only terms exponential in N will retain statistical relevance for $N \to \infty$:

$$\begin{aligned}\lim_{N\to\infty} \frac{1}{N} \log \mathcal{D}(\mathbf{m}) &= \lim_{N\to\infty} \frac{1}{N} \log \int d\mathbf{x}\ \mathrm{e}^{iN\mathbf{x}\cdot\mathbf{m}} \langle \mathrm{e}^{-i\sigma_i \boldsymbol{\xi}_i \cdot \mathbf{x}} \rangle_{\boldsymbol{\sigma}} \\ &= \lim_{N\to\infty} \frac{1}{N} \log \int d\mathbf{x}\ \mathrm{e}^{N[i\mathbf{x}\cdot\mathbf{m} + \langle \log\cos[\boldsymbol{\xi}\cdot\mathbf{x}]\rangle_{\boldsymbol{\xi}}]},\end{aligned}$$

with the abbreviation $\langle \Phi(\boldsymbol{\xi}) \rangle_{\boldsymbol{\xi}} = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \Phi(\boldsymbol{\xi}_i)$. The leading contribution to both free energies can be expressed as a finite-dimensional integral, for large N dominated by that saddle–point (extremum) for which the extensive exponent is real and maximal [Coo01, SC00, SC01, CKS05]:

$$\begin{aligned}&\lim_{N\to\infty} F/N = -\frac{1}{\beta N} \log \int d\mathbf{m} d\mathbf{x}\ \mathrm{e}^{-N\beta f(\mathbf{m},\mathbf{x})} = \mathrm{extr}_{\mathbf{x},\mathbf{m}}\ f(\mathbf{m},\mathbf{x}), \\ &\lim_{N\to\infty} \tilde{F}/N = -\frac{1}{\beta N} \log \int d\mathbf{m} d\mathbf{x}\ \mathrm{e}^{-N\beta \tilde{f}(\mathbf{m},\mathbf{x})} = \mathrm{extr}_{\mathbf{x},\mathbf{m}}\ \tilde{f}(\mathbf{m},\mathbf{x}), \qquad \text{with}\end{aligned}$$

$$f(\mathbf{m},\mathbf{x}) = -\tfrac{1}{2}\mathbf{m}^2 - i\mathbf{x}\cdot\mathbf{m} - \beta^{-1} \langle \log 2\cos[\beta\boldsymbol{\xi}\cdot\mathbf{x}] \rangle_{\boldsymbol{\xi}},$$

$$\tilde{f}(\mathbf{m},\mathbf{x}) = -\beta^{-1} \langle \log 2\cosh[\beta\boldsymbol{\xi}\cdot\mathbf{m}] \rangle_{\boldsymbol{\xi}} - i\mathbf{x}\cdot\mathbf{m} - \beta^{-1} \langle \log 2\cos[\beta\boldsymbol{\xi}\cdot\mathbf{x}] \rangle_{\boldsymbol{\xi}}.$$

The saddle–point equations for f and $\tilde{f}$ are given by:

$$\begin{aligned}&f : \mathbf{x} = i\mathbf{m}, && i\mathbf{m} = \langle \boldsymbol{\xi} \tan[\beta\boldsymbol{\xi}\cdot\mathbf{x}] \rangle_{\boldsymbol{\xi}}, \\ &\tilde{f} : \mathbf{x} = i\langle \boldsymbol{\xi} \tanh[\beta\boldsymbol{\xi}\cdot\mathbf{m}] \rangle_{\boldsymbol{\xi}}, && i\mathbf{m} = \langle \boldsymbol{\xi} \tan[\beta\boldsymbol{\xi}\cdot\mathbf{x}] \rangle_{\boldsymbol{\xi}}.\end{aligned}$$

In saddle-points $\mathbf{x}$ turns out to be purely imaginary. However, after a shift of the integration contours, putting $\mathbf{x} = i\mathbf{x}^\star(\mathbf{m}) + \mathbf{y}$ (where $i\mathbf{x}^\star(\mathbf{m})$ is the imaginary saddle–point, and where $\mathbf{y} \in \mathbb{R}^p$) we can eliminate $\mathbf{x}$ in favor of $\mathbf{y} \in \mathbb{R}^p$ which does have a real saddle–point, by construction.(Our functions to be integrated have no poles, but strictly speaking we still have to verify that the integration segments linking the original integration regime to the shifted one will not contribute to the integrals. This is generally a tedious and distracting task, which is often skipped. For simple models, however (e.g., networks with uniform synapses), the verification can be carried out properly, and all is found to be safe.) We then get

$$\begin{aligned}&\text{Sequential Dynamics: } \mathbf{m} = \langle \boldsymbol{\xi} \tanh[\beta\boldsymbol{\xi}\cdot\mathbf{m}] \rangle_{\boldsymbol{\xi}}, \\ &\text{Parallel Dynamics: } \quad \mathbf{m} = \langle \boldsymbol{\xi} \tanh[\beta\boldsymbol{\xi}\cdot[\langle \boldsymbol{\xi}' \tanh[\beta\boldsymbol{\xi}'\cdot\mathbf{m}] \rangle_{\boldsymbol{\xi}'}]] \rangle_{\boldsymbol{\xi}},\end{aligned}$$

(compare to e.g., (2.157,2.158)). The solutions of the above two equations will in general be identical. To see this, let us denote $\hat{\mathbf{m}} = \langle \boldsymbol{\xi} \tanh[\beta\boldsymbol{\xi}\cdot\mathbf{m}] \rangle_{\boldsymbol{\xi}}$, with which the saddle point equation for $\tilde{f}$ decouples into:

$$\mathbf{m} = \langle \boldsymbol{\xi} \tanh\left[\beta \boldsymbol{\xi} \cdot \hat{\mathbf{m}}\right] \rangle_{\boldsymbol{\xi}}, \qquad \hat{\mathbf{m}} = \langle \boldsymbol{\xi} \tanh\left[\beta \boldsymbol{\xi} \cdot \mathbf{m}\right] \rangle_{\boldsymbol{\xi}}, \qquad \text{so}$$
$$\left[\mathbf{m} - \hat{\mathbf{m}}\right]^2 = \langle \left[(\boldsymbol{\xi} \cdot \mathbf{m}) - (\boldsymbol{\xi} \cdot \hat{\mathbf{m}})\right] \left[\tanh(\beta \boldsymbol{\xi} \cdot \hat{\mathbf{m}}) - \tanh(\beta \boldsymbol{\xi} \cdot \mathbf{m})\right] \rangle_{\boldsymbol{\xi}}.$$

Since tanh is a monotonicaly–increasing function, we must have $[\mathbf{m} - \hat{\mathbf{m}}] \cdot \boldsymbol{\xi} = 0$ for each $\boldsymbol{\xi}$ that contributes to the averages $\langle \ldots \rangle_{\boldsymbol{\xi}}$. For all choices of patterns where the covariance matrix $C_{\mu\nu} = \langle \xi_\mu \xi_\nu \rangle_{\boldsymbol{\xi}}$ is positive definite, we thus get $\mathbf{m} = \hat{\mathbf{m}}$. The final result is: for both types of dynamics (sequential and parallel) the overlap order parameters in equilibrium are given by the solution $\mathbf{m}^*$ of

$$\mathbf{m} = \langle \boldsymbol{\xi} \tanh\left[\beta \boldsymbol{\xi} \cdot \mathbf{m}\right] \rangle_{\boldsymbol{\xi}}, \qquad \text{which minimises} \tag{2.164}$$

$$f(\mathbf{m}) = \frac{1}{2}\mathbf{m}^2 - \frac{1}{\beta} \langle \log 2 \cosh\left[\beta \boldsymbol{\xi} \cdot \mathbf{m}\right] \rangle_{\boldsymbol{\xi}}. \tag{2.165}$$

The free energies of the ergodic components are $\lim_{N\to\infty} F/N = f(\mathbf{m}^*)$ and $\lim_{N\to\infty} \tilde{F}/N = 2f(\mathbf{m}^*)$. Adding partition terms of the form $H \to H + \lambda g[\mathbf{m}(\boldsymbol{\sigma})]$ to the Hamiltonians allows us identify $\langle g[\mathbf{m}(\boldsymbol{\sigma})] \rangle_{\rm eq} = \lim_{\lambda\to 0} \partial F/\partial\lambda = g[\mathbf{m}^*]$. Thus, in equilibrium the fluctuations in the overlap order parameters $\mathbf{m}(\boldsymbol{\sigma})$ (2.160) vanish for $N \to \infty$. Their deterministic values are simply given by $\mathbf{m}^*$. Note that in the case of sequential dynamics we could also have used linearization with Gaussian integrals (as used previously for coupled oscillators with uniform synapses) to arrive at this solution, with p auxiliary integrations, but that for parallel dynamics this would not have been possible.

Now, in analysis of *order parameter equations*, we will restrict our further discussion to the case of randomly drawn patterns, so [Coo01, SC00, SC01, CKS05]

$$\langle \Phi(\boldsymbol{\xi}) \rangle_{\boldsymbol{\xi}} = 2^{-p} \sum_{\boldsymbol{\xi} \in \{-1,1\}^p} \Phi(\boldsymbol{\xi}), \qquad \langle \xi_\mu \rangle_{\boldsymbol{\xi}} = 0, \qquad \langle \xi_\mu \xi_\nu \rangle_{\boldsymbol{\xi}} = \delta_{\mu\nu}.$$

We first establish an upper bound for the temperature for where non–trivial solutions $\mathbf{m}^*$ could exist, by writing (2.164) in integral form:

$$m_\mu = \beta \langle \xi_\mu (\boldsymbol{\xi} \cdot \mathbf{m}) \int_0^1 d\lambda [1 - \tanh^2[\beta\lambda \boldsymbol{\xi} \cdot \mathbf{m}]] \rangle_{\boldsymbol{\xi}},$$

from which we deduce

$$0 = \mathbf{m}^2 - \beta \langle (\boldsymbol{\xi} \cdot \mathbf{m})^2 \int_0^1 d\lambda [1 - \tanh^2\left[\beta\lambda \boldsymbol{\xi} \cdot \mathbf{m}\right]] \rangle_{\boldsymbol{\xi}}$$
$$\geq \mathbf{m}^2 - \beta \langle (\boldsymbol{\xi} \cdot \mathbf{m})^2 \rangle_{\boldsymbol{\xi}} = \mathbf{m}^2 (1 - \beta),$$

For $T > 1$ the only solution of (2.164) is the paramagnetic state $\mathbf{m} = 0$, which gives for the free energy per neuron $-T \log 2$ and $-2T \log 2$ (for sequential and parallel dynamics, respectively). At $T = 1$ a phase transition occurs, which follows from expanding (2.164) for small $|\mathbf{m}|$ in powers of $\tau = \beta - 1$:

$$m_\mu = (1+\tau)m_\mu - \frac{1}{3}m_\nu m_\rho m_\lambda \langle \xi_\mu \xi_\nu \xi_\rho \xi_\lambda \rangle_{\boldsymbol{\xi}}$$
$$+ \mathcal{O}(\mathbf{m}^5, \tau \mathbf{m}^3) = m_\mu [1 + \tau - \mathbf{m}^2 + \frac{2}{3} m_\mu^2] + \mathcal{O}(\mathbf{m}^5, \tau \mathbf{m}^3).$$

The new saddle–point scales as $m_\mu = \tilde{m}_\mu \tau^{1/2} + \mathcal{O}(\tau^{3/2})$, with for each μ: $\tilde{m}_\mu = 0$ or $0 = 1 - \tilde{\mathbf{m}}^2 + \frac{2}{3}\tilde{m}_\mu^2$.
The solutions are of the form $\tilde{m}_\mu \in \{-\tilde{m}, 0, \tilde{m}\}$. If we denote with n the number of non-zero components in the vector $\tilde{\mathbf{m}}$, we derive from the above identities: $\tilde{m}_\mu = 0$ or $\tilde{m}_\mu = \pm\sqrt{3}/\sqrt{3n-2}$. These saddle-points are called *mixture states*, since they correspond to microscopic configurations correlated equally with a finite number n of the stored patterns (or their negatives). Without loss of generality we can always perform gauge transformations on the set of stored patterns (permutations and reflections), such that the mixture states acquire the form [Coo01, SC00, SC01, CKS05]

$$\mathbf{m} = m_n(\overbrace{1, \ldots, 1}^{n \text{ times}}, \overbrace{0, \ldots, 0}^{p-n \text{ times}}), \qquad (2.166)$$
$$m_n = [\frac{3}{3n-2}]^{\frac{1}{2}} (\beta - 1)^{1/2} + \ldots$$

These states are in fact saddle–points of the surface $f(\mathbf{m})$ (2.165) for any finite temperature, as can be verified by substituting (2.166) as an *ansatz* into (2.164):

$$\mu \leq n: \qquad m_n = \langle \xi_\mu \tanh[\beta m_n \sum_{\nu \leq n} \xi_\nu] \rangle_{\boldsymbol{\xi}},$$
$$\mu > n: \qquad 0 = \langle \xi_\mu \tanh[\beta m_n \sum_{\nu \leq n} \xi_\nu] \rangle_{\boldsymbol{\xi}}.$$

The second equation is automatically satisfied since the average factorizes. The first equation leads to a condition determining the amplitude m_n of the mixture states:

$$m_n = \langle [\frac{1}{n} \sum_{\mu \leq n} \xi_\mu] \tanh[\beta m_n \sum_{\nu \leq n} \xi_\nu] \rangle_{\boldsymbol{\xi}}. \qquad (2.167)$$

The corresponding values of $f(\mathbf{m})$, to be denoted by f_n, are

$$f_n = \frac{1}{2} n m_n^2 - \frac{1}{\beta} \langle \log 2 \cosh[\beta m_n \sum_{\nu \leq n} \xi_\nu] \rangle_{\boldsymbol{\xi}}. \qquad (2.168)$$

The relevant question at this stage is whether or not these saddle-points correspond to local minima of the surface $f(\mathbf{m})$ (2.165). The second derivative of $f(\mathbf{m})$ is given by

$$\frac{\partial^2 f(\mathbf{m})}{\partial m_\mu \partial m_\nu} = \delta_{\mu\nu} - \beta \langle \xi_\mu \xi_\nu \left[1 - \tanh^2 [\beta \boldsymbol{\xi} \cdot \mathbf{m}]\right] \rangle_{\boldsymbol{\xi}}, \qquad (2.169)$$

where a local minimum corresponds to a positive definite second derivative. In the trivial saddle–point $\mathbf{m} = 0$ this gives simply $\delta_{\mu\nu}(1-\beta)$, so at $T = 1$ this state destabilizes. In a mixture state of the type (2.166) the second derivative becomes:

$$D^{(n)}_{\mu\nu} = \delta_{\mu\nu} - \beta\langle\xi_\mu\xi_\nu[1 - \tanh^2[\beta m_n \sum_{\rho\leq n} \xi_\rho]]\rangle_{\boldsymbol{\xi}}.$$

Due to the symmetries in the problem the spectrum of the matrix $D^{(n)}$ can be calculated. One finds the following eigen–spaces, with

$$Q = \langle\tanh^2[\beta m_n \sum_{\rho\leq n} \xi_\rho]\rangle_{\boldsymbol{\xi}} \qquad \text{and} \qquad R = \langle\xi_1\xi_2 \tanh^2[\beta m_n \sum_{\rho\leq n} \xi_\rho]\rangle_{\boldsymbol{\xi}},$$

	Eigenspace :	Eigenvalue :
I :	$\mathbf{x} = (0, \ldots, 0, x_{n+1}, \ldots, x_p)$,	$1 - \beta[1 - Q]$,
II :	$\mathbf{x} = (1, \ldots, 1, 0, \ldots, 0)$,	$1 - \beta[1 - Q + (1 - n)R]$,
III :	$\mathbf{x} = (x_1, \ldots, x_n, 0, \ldots, 0)$, $\sum_\mu x_\mu = 0$,	$1 - \beta[1 - Q + R]$.

The eigen–space III and the quantity R only come into play for $n > 1$. To find the smallest eigenvalue we need to know the sign of R. With the abbreviation $M_{\boldsymbol{\xi}} = \sum_{\rho\leq n} \xi_\rho$ we find [Coo01, SC00, SC01, CKS05]:

$$\begin{aligned} n(n-1)R &= \langle M^2_{\boldsymbol{\xi}} \tanh^2[\beta m_n M_{\boldsymbol{\xi}}]\rangle_{\boldsymbol{\xi}} - n\langle\tanh^2[\beta m_n M_{\boldsymbol{\xi}}]\rangle_{\boldsymbol{\xi}} \\ &= \langle[M^2_{\boldsymbol{\xi}} - \langle M^2_{\boldsymbol{\xi}'}\rangle_{\boldsymbol{\xi}'}] \tanh^2[\beta m_n |M_{\boldsymbol{\xi}}|]\rangle_{\boldsymbol{\xi}} \\ &= \langle[M^2_{\boldsymbol{\xi}} - \langle M^2_{\boldsymbol{\xi}'}\rangle_{\boldsymbol{\xi}'}] \left\{\tanh^2[\beta m_n \sqrt{M^2_{\boldsymbol{\xi}}}] - \tanh^2[\beta m_n \sqrt{\langle M^2_{\boldsymbol{\xi}'}\rangle_{\boldsymbol{\xi}'}}]\right\}\rangle_{\boldsymbol{\xi}} \geq 0. \end{aligned}$$

We may now identify the conditions for an $n-$mixture state to be a local minimum of $f(\mathbf{m})$. For $n = 1$ the relevant eigenvalue is I, now the quantity Q simplifies considerably. For $n > 1$ the relevant eigenvalue is III, here we can combine Q and R into one single average:

$$\begin{aligned} &n = 1 : 1 - \beta[1 - \tanh^2[\beta m_1]] > 0 \\ &n = 2 : 1 - \beta > 0 \\ &n \geq 3 : 1 - \beta[1 - \langle\tanh^2[\beta m_n \textstyle\sum_{\rho=3}^n \xi_\rho]\rangle_{\boldsymbol{\xi}}] > 0 \end{aligned}$$

The $n = 1$ states, correlated with one pattern only, are the desired solutions. They are stable for all $T < 1$, since partial differentiation with respect to β of the $n = 1$ amplitude equation (2.167) gives

$$\begin{aligned} m_1 = \tanh[\beta m_1] \quad &\rightarrow \quad 1 - \beta[1 - \tanh^2[\beta m_1]] \\ &= m_1[1 - \tanh^2[\beta m_1]](\partial m_1/\partial\beta)^{-1}, \end{aligned}$$

so that clearly $\mathrm{sgn}[m_1] = \mathrm{sgn}[\partial m_1/\partial\beta]$. The $n = 2$ mixtures are always unstable. For $n \geq 3$ we have to solve the amplitude equations (2.167) numerically to evaluate their stability. It turns out that only for odd n will there be a

critical temperature below which the $n-$mixture states are local minima of $f(\mathbf{m})$.

We have now solved the model in equilibrium for finite p and $N \to \infty$. For non–random patterns one simply has to study the bifurcation properties of equation (2.164) for the new pattern statistics at hand; this is only qualitatively different from the random pattern analysis explained above. The occurrence of multiple saddle–points corresponding to local minima of the free energy signals ergodicity breaking. Although among these only the *global minimum* will correspond to the thermodynamic equilibrium state, the non–global minima correspond to true ergodic components, i.e., on finite time–scales they will be just as relevant as the global minimum.

2.3.5 Simple Recurrent Networks of Coupled Oscillators

Coupled Oscillators with Uniform Synapses

Models with continuous variables involve integration over states, rather than summation. For a coupled oscillator network (2.133) with uniform synapses $J_{ij} = J/N$ and zero frequencies $\omega_i = 0$ (which is a simple version of the model in [Kur84]) we get for the free energy per oscillator [Coo01, SC00, SC01, CKS05]:

$$\lim_{N\to\infty} F/N = -\lim_{N\to\infty} \frac{1}{\beta N} \log \int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} d\boldsymbol{\xi} \ \times \\ \times \mathrm{e}^{(\beta J/2N)\left[\left[\sum_i \cos(\xi_i)\right]^2 + \left[\sum_i \sin(\xi_i)\right]^2\right]}.$$

We would now have to 'count' microscopic states with prescribed average cosines and sines. A faster route exploits auxiliary Gaussian integrals, via the identity

$$\mathrm{e}^{\frac{1}{2}y^2} = \int Dz \ \mathrm{e}^{yz}, \tag{2.170}$$

with the short–hand $Dx = (2\pi)^{-\frac{1}{2}} \mathrm{e}^{-\frac{1}{2}x^2} dx$ (this alternative would also have been open to us in the binary case; my aim in this section is to explain both methods):

$$\begin{aligned}
\lim_{N\to\infty} F/N &= -\lim_{N\to\infty} \frac{1}{\beta N} \log \int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} d\boldsymbol{\xi} \int Dx Dy \ \times \\
&\quad \times \mathrm{e}^{\sqrt{\beta J/N}\left[x \sum_i \cos(\xi_i) + y \sum_i \sin(\xi_i)\right]} \\
&= -\lim_{N\to\infty} \frac{1}{\beta N} \log \int Dx Dy \left[\int_{-\pi}^{\pi} d\xi \ \mathrm{e}^{\cos(\xi)\sqrt{\beta J(x^2+y^2)/N}}\right]^N \\
&= -\lim_{N\to\infty} \frac{1}{\beta N} \log \int_0^{\infty} dq \ q e^{-\frac{1}{2}N\beta|J|q^2} \times \\
&\quad \times \left[\int_{-\pi}^{\pi} d\xi \ \mathrm{e}^{\beta|J|q\cos(\xi)\sqrt{\mathrm{rmsgn}(J)}}\right]^N,
\end{aligned}$$

where we have transformed to polar coordinates, $(x, y) = q\sqrt{\beta|J|N}(\cos\theta, \sin\theta)$, and where we have already eliminated (constant) terms which will not survive the limit $N \to \infty$. Thus, saddle–point integration gives us, quite similar to the previous cases (2.155,2.156):

$$\lim_{N\to\infty} F/N = \min_{q\geq 0} f(q), \qquad \begin{array}{l} J > 0 : \beta f(q) = \frac{1}{2}\beta|J|q^2 - \log[2\pi I_0(\beta|J|q)] \\ J < 0 : \beta f(q) = \frac{1}{2}\beta|J|q^2 - \log[2\pi I_0(i\beta|J|q)] \end{array}, \tag{2.171}$$

in which the $I_n(z)$ are the Bessel functions (see e.g., [AS72]). The equations from which to solve the minima are got by differentiation, using $\frac{d}{dz}I_0(z) = I_1(z)$:in which the $I_n(z)$ are the *Bessel functions* (see e.g., [AS72]). The equations from which to solve the minima are got by differentiation, using $\frac{d}{dz}I_0(z) = I_1(z)$:

$$J > 0 : \qquad q = \frac{I_1(\beta|J|q)}{I_0(\beta|J|q)}, \qquad J < 0 : \qquad q = i\,\frac{I_1(i\beta|J|q)}{I_0(i\beta|J|q)}. \tag{2.172}$$

Again, in both cases the problem has been reduced to studying a single non-linear equation. The physical meaning of the solution follows from the identity $-2\partial F/\partial J = \langle N^{-1}\sum_{i\neq j}\cos(\xi_i - \xi_j)\rangle$:

$$\lim_{N\to\infty}\langle[\frac{1}{N}\sum_i \cos(\xi_i)]^2\rangle + \lim_{N\to\infty}\langle[\frac{1}{N}\sum_i \sin(\xi_i)]^2\rangle = \mathrm{sgn}(J)\ q^2.$$

From this equation it also follows that $q \leq 1$. Note: since $\partial f(q)/\partial q = 0$ at the minimum, one only needs to consider the explicit derivative of $f(q)$ with respect to J. If the synapses induce anti-synchronization, $J < 0$, the only solution of (2.172) (and the minimum in (2.171)) is the trivial state $q = 0$. This also follows immediately from the equation which gave the physical meaning of q. For synchronizing forces, $J > 0$, on the other hand, we again find the trivial solution at high noise levels, but a globally synchronized state with $q > 0$ at low noise levels. Here a phase transition occurs at $T = \frac{1}{2}J$ (a bifurcation of non–trivial solutions of (2.172)), and for $T < \frac{1}{2}J$ the minimum of (2.171) is found at two non-zero values for q. The critical noise level is again found upon expanding the saddle–point equation, using

$$I_0(z) = 1 + \mathcal{O}(z^2) \qquad \text{and} \qquad I_1(z) = \frac{1}{2}z + \mathcal{O}(z^3) : q = \frac{1}{2}\beta J q + \mathcal{O}(q^3).$$

Precisely at $\beta J = 2$ one finds a de-stabilization of the trivial solution $q = 0$, together with the creation of (two) stable non–trivial ones. Note that, in view of (2.171), we are only interested in non–negative values of q. One can prove, using the properties of the Bessel functions, that there are no other (discontinuous) bifurcations of non–trivial solutions of the saddle–point equation. Note, finally, that the absence of a state with global anti-synchronization for

$J < 0$ has the same origin as the absence of an anti-ferromagnetic state for $J < 0$ in the previous models with binary neurons. Due to the long-range nature of the synapses $J_{ij} = J/N$ such states simply cannot exist: whereas any set of oscillators can be in a fully synchronized state, if two oscillators are in anti-synchrony it is already impossible for a third to be simultaneously in anti-synchrony with the first two (since anti-synchrony with one implies synchrony with the other) [Coo01, SC00, SC01, CKS05].

Coupled Oscillator Attractor Networks

Let us now turn to an alternative realisation of information storage in a recurrent network based upon the creation of attractors [Coo01, SC00, SC01, CKS05]. We will solve models of coupled neural oscillators of the type (2.133), with zero natural frequencies (since we wish to use equilibrium techniques), in which real-valued patterns are stored as stable configurations of oscillator phases, following [Coo89]. Let us, however, first find out how to store a single pattern $\boldsymbol{\xi} \in [-\pi, \pi]^N$ in a noise-less infinite-range oscillator network. For simplicity we will draw each component ξ_i independently at random from $[-\pi, \pi]$, with uniform probability density. This allows us to use asymptotic properties such as $|N^{-1} \sum_j \mathrm{e}^{i\ell\xi_j}| = \mathcal{O}(N^{-\frac{1}{2}})$ for any integer ℓ. A sensible choice for the synapses would be $J_{ij} = \cos[\xi_i - \xi_j]$. To see this we work out the corresponding Lyapunov function (2.140):

$$L[\boldsymbol{\xi}] = -\frac{1}{2N^2} \cos[\xi_i - \xi_j] \cos[\xi_i - \xi_j],$$
$$L[\boldsymbol{\xi}] = -\frac{1}{2N^2} \cos^2[\xi_i - \xi_j] = -\frac{1}{4} + \mathcal{O}(N^{-\frac{1}{2}}),$$

where the factors of N have been inserted to achieve appropriate scaling in the $N \to \infty$ limit. The function $L[\boldsymbol{\xi}]$, which is obviously bounded from below, must decrease monotonically during the dynamics. To find out whether the state $\boldsymbol{\xi}$ is a stable fixed–point of the dynamics we have to calculate L and derivatives of L at $\boldsymbol{\xi} = \boldsymbol{\xi}$:

$$\left.\frac{\partial L}{\partial \xi_i}\right|_{\boldsymbol{\xi}} = \frac{1}{2N^2} \sum_j \sin[2(\xi_i - \xi_j)],$$
$$\left.\frac{\partial^2 L}{\partial \xi_i^2}\right|_{\boldsymbol{\xi}} = \frac{1}{N^2} \sum_j \cos^2[\xi_i - \xi_j],$$
$$i \neq j: \quad \left.\frac{\partial^2 L}{\partial \xi_i \partial \xi_j}\right|_{\boldsymbol{\xi}} = -\frac{1}{N^2} \cos^2[\xi_i - \xi_j].$$

Clearly $\lim_{N\to\infty} L[\boldsymbol{\xi}] = -\frac{1}{4}$. Putting $\boldsymbol{\xi} = \boldsymbol{\xi} + \Delta\boldsymbol{\xi}$, with $\Delta\xi_i = \mathcal{O}(N^0)$, we find

$$L[\boldsymbol{\xi}+\Delta\boldsymbol{\xi}]-L[\boldsymbol{\xi}]=\Delta\xi_i\frac{\partial L}{\partial\xi_i}|_{\boldsymbol{\xi}} \tag{2.173}$$
$$+\frac{1}{2}\Delta\xi_i\Delta\xi_j\frac{\partial^2 L}{\partial\xi_i\partial\xi_j}|_{\boldsymbol{\xi}}+\mathcal{O}(\Delta\boldsymbol{\xi}^3)$$
$$=\frac{1}{4N}\sum_i\Delta\xi_i^2-\frac{1}{2N^2}\Delta\xi_i\Delta\xi_j\cos^2[\xi_i-\xi_j]+\mathcal{O}(N^{-\frac{1}{2}},\Delta\boldsymbol{\xi}^3)$$

$$=\frac{1}{4}\{\frac{1}{N}\sum_i\Delta\xi_i^2-[\frac{1}{N}\sum_i\Delta\xi_i]^2-[\frac{1}{N}\Delta\xi_i\cos(2\xi_i)]^2$$
$$-[\frac{1}{N}\sum_i\Delta\xi_i\sin(2\xi_i)]^2\}+\mathcal{O}(N^{-\frac{1}{2}},\Delta\boldsymbol{\xi}^3).$$

In leading order in N the following three vectors in $\mathbb{R}^N$ are normalized and orthogonal:

$$\mathbf{e}_1=\frac{1}{\sqrt{N}}(1,1,\ldots,1),\qquad \mathbf{e}_2=\frac{\sqrt{2}}{\sqrt{N}}(\cos(2\xi_1),\ldots,\cos(2\xi_N)),$$
$$\mathbf{e}_2=\frac{\sqrt{2}}{\sqrt{N}}(\sin(2\xi_1),\ldots,\sin(2\xi_N)).$$

We may therefore use

$$\Delta\boldsymbol{\xi}^2\geq(\Delta\boldsymbol{\xi}\cdot_1)^2+(\Delta\boldsymbol{\xi}\cdot_2)^2+(\Delta\boldsymbol{\xi}\cdot_3)^2,$$

insertion of which into (2.173) leads to

$$L[\boldsymbol{\xi}+\Delta\boldsymbol{\xi}]-L[\boldsymbol{\xi}]\ \geq\ [\frac{1}{2N}\sum_i\Delta\xi_i\cos(2\xi_i)]^2$$
$$+[\frac{1}{2N}\sum_i\Delta\xi_i\sin(2\xi_i)]^2+\mathcal{O}(N^{-\frac{1}{2}},\Delta\boldsymbol{\xi}^3).$$

Thus for large N the second derivative of L is non-negative at $\boldsymbol{\xi}=\boldsymbol{\xi}$, and the phase pattern $\boldsymbol{\xi}$ has indeed become a fixed–point attractor of the dynamics of the noise-free coupled oscillator network. The same is found to be true for the states $\boldsymbol{\xi}=\pm\boldsymbol{\xi}+\alpha(1,\ldots,1)$ (for any α).

We next follow the strategy of the Hopfield model and attempt to simply extend the above recipe for the synapses to the case of having a finite number p of phase patterns $\boldsymbol{\xi}^\mu=(\xi_1^\mu,\ldots,\xi_N^\mu)\in[-\pi,\pi]^N$, giving

$$J_{ij}=\frac{1}{N}\sum_{\mu=1}^{p}\cos[\xi_i^\mu-\xi_j^\mu], \tag{2.174}$$

where the factor N, as before, ensures a proper limit $N\to\infty$ later. In analogy with our solution of the Hopfield model we define the following averages over pattern variables:

$$\langle g[\boldsymbol{\xi}]\rangle_{\boldsymbol{\xi}} = \lim_{N\to\infty}\sum_i g[\boldsymbol{\xi}_i], \qquad \boldsymbol{\xi}_i = (\xi_i^1,\ldots,\xi_i^p)\in[-\pi,\pi]^p.$$

We can write the Hamiltonian $H(\boldsymbol{\xi})$ of (2.153) in the form [Coo01, SC00, SC01, CKS05]

$$\begin{aligned} H(\boldsymbol{\xi}) &= -\frac{1}{2N}\sum_{\mu=1}^{p}\cos[\xi_i^\mu-\xi_j^\mu]\cos[\xi_i-\xi_j] \\ &= -\frac{N}{2}\sum_{\mu=1}^{p}\left\{m_{cc}^\mu(\boldsymbol{\xi})^2+m_{cs}^\mu(\boldsymbol{\xi})^2+m_{sc}^\mu(\boldsymbol{\xi})^2+m_{ss}^\mu(\boldsymbol{\xi})^2\right\}, \end{aligned}$$

in which

$$\begin{aligned} m_{cc}^\mu(\boldsymbol{\xi}) &= \frac{1}{N}\cos(\xi_i^\mu)\cos(\xi_i), \\ m_{cs}^\mu(\boldsymbol{\xi}) &= \frac{1}{N}\cos(\xi_i^\mu)\sin(\xi_i), \end{aligned} \tag{2.175}$$

$$\begin{aligned} m_{sc}^\mu(\boldsymbol{\xi}) &= \frac{1}{N}\sin(\xi_i^\mu)\cos(\xi_i), \\ m_{ss}^\mu(\boldsymbol{\xi}) &= \frac{1}{N}\sin(\xi_i^\mu)\sin(\xi_i). \end{aligned} \tag{2.176}$$

The free energy per oscillator can now be written as

$$\begin{aligned} F/N = &-\frac{1}{\beta N}\log\int\cdots\int d\boldsymbol{\xi}\ \mathrm{e}^{-\beta H(\boldsymbol{\xi})} = \\ &-\frac{1}{\beta N}\log\int\cdots\int d\boldsymbol{\xi}\ \mathrm{e}^{\frac{1}{2}\beta N\sum_\mu\sum_{\star\star}m_{\star\star}^\mu(\boldsymbol{\xi})^2}, \end{aligned}$$

with $\star\star\in\{cc,ss,cs,sc\}$. Upon introducing the notation $\mathbf{m}_{\star\star}=(m_{\star\star}^1,\ldots,m_{\star\star}^p)$ we can again express the free energy in terms of the density of states $\mathcal{D}(\{\mathbf{m}_{\star\star}\})=(2\pi)^{-N}\int\cdots\int d\boldsymbol{\xi}\ \prod_{\star\star}\delta[\mathbf{m}_{\star\star}-\mathbf{m}_{\star\star}(\boldsymbol{\sigma})]$:

$$F/N = -\frac{1}{\beta}\log(2\pi)-\frac{1}{\beta N}\log\int\prod_{\star\star}d\mathbf{m}_{\star\star}\ \mathcal{D}(\{\mathbf{m}_{\star\star}\})\mathrm{e}^{\frac{1}{2}\beta N\sum_{\star\star}\mathbf{m}_{\star\star}^2}. \tag{2.177}$$

Since p is finite, the leading contribution to the density of states (as $N\to\infty$), which will give us the entropy, can be calculated by writing the $\delta-$functions in integral representation:

$$\lim_{N\to\infty}\frac{1}{N}\log\mathcal{D}(\{\mathbf{m}_{\star\star}\})=\lim_{N\to\infty}\frac{1}{N}\log\int\prod_{\star\star}\left[d\mathbf{x}_{\star\star}\ \mathrm{e}^{iN\mathbf{x}_{\star\star}\cdot\mathbf{m}_{\star\star}}\right]\ \times$$

$$\int\cdots\int\frac{d\boldsymbol{\xi}}{(2\pi)^N}\ \times$$

$$\mathrm{e}^{-i[x^{\mu}_{cc}\cos(\xi^{\mu}_i)\cos(\xi_i)+x^{\mu}_{cs}\cos(\xi^{\mu}_i)\sin(\xi_i)+x^{\mu}_{sc}\sin(\xi^{\mu}_i)\cos(\xi_i)+x^{\mu}_{ss}\sin(\xi^{\mu}_i)\sin(\xi_i)]}$$

$$=\mathrm{extr}_{\{\mathbf{x}_{\star\star}\}}\{i\sum_{\star\star}\mathbf{x}_{\star\star}\cdot\mathbf{m}_{\star\star}+\langle\log\int\frac{d\xi}{2\pi}\times$$

$$\mathrm{e}^{-i[x^{\mu}_{cc}\cos(\xi_\mu)\cos(\xi)+x^{\mu}_{cs}\cos(\xi_\mu)\sin(\xi)+x^{\mu}_{sc}\sin(\xi_\mu)\cos(\xi)+x^{\mu}_{ss}\sin(\xi_\mu)\sin(\xi)]}\rangle_{\boldsymbol{\xi}}\}.$$

The relevant extremum is purely imaginary so we put $\mathbf{x}_{\star\star}=i\beta\mathbf{y}_{\star\star}$ (see our previous discussion for the Hopfield model) and, upon inserting the density of states into our original expression for the free energy per oscillator, arrive at

$$\lim_{N\to\infty}F/N=\mathrm{extr}_{\{\mathbf{m}_{\star\star},\mathbf{y}_{\star\star}\}}\ f(\{\mathbf{m}_{\star\star},\mathbf{y}_{\star\star}\}),$$

$$f(\{\mathbf{m}_{\star\star},\mathbf{y}_{\star\star}\})=-\frac{1}{\beta}\log(2\pi)-\frac{1}{2}\sum_{\star\star}\mathbf{m}^2_{\star\star}+\sum_{\star\star}\mathbf{y}_{\star\star}\cdot\mathbf{m}_{\star\star}$$

$$-\frac{1}{\beta}\langle\log\int\frac{d\xi}{2\pi}\mathrm{e}^{\beta[y^{\mu}_{cc}\cos(\xi_\mu)\cos(\xi)+y^{\mu}_{cs}\cos(\xi_\mu)\sin(\xi)+y^{\mu}_{sc}\sin(\xi_\mu)\cos(\xi)+y^{\mu}_{ss}\sin(\xi_\mu)\sin(\xi)]}\rangle_{\boldsymbol{\xi}}$$

Taking derivatives with respect to the order parameters $\mathbf{m}_{\star\star}$ gives us $\mathbf{y}_{\star\star}=\mathbf{m}_{\star\star}$, with which we can eliminate the $\mathbf{y}_{\star\star}$. Derivation with respect to the $\mathbf{m}_{\star\star}$ subsequently gives the saddle–point equations

$$m^{\mu}_{cc}= \tag{2.178}$$

$$\langle\cos[\xi_\mu]\frac{\int d\xi\cos[\xi]\mathrm{e}^{\beta\cos[\xi][m^{\nu}_{cc}\cos[\xi_\nu]+m^{\nu}_{sc}\sin[\xi_\nu]]+\beta\sin[\xi][m^{\nu}_{cs}\cos[\xi_\nu]+m^{\nu}_{ss}\sin[\xi_\nu]]}}{\int d\xi\ \mathrm{e}^{\beta\cos[\xi][m^{\nu}_{cc}\cos[\xi_\nu]+m^{\nu}_{sc}\sin[\xi_\nu]]+\beta\sin[\xi][m^{\nu}_{cs}\cos[\xi_\nu]+m^{\nu}_{ss}\sin[\xi_\nu]]}}\rangle_{\boldsymbol{\xi}},$$

$$m^{\mu}_{cs}= \tag{2.179}$$

$$\langle\cos[\xi_\mu]\frac{\int d\xi\sin[\xi]\mathrm{e}^{\beta\cos[\xi][m^{\nu}_{cc}\cos[\xi_\nu]+m^{\nu}_{sc}\sin[\xi_\nu]]+\beta\sin[\xi][m^{\nu}_{cs}\cos[\xi_\nu]+m^{\nu}_{ss}\sin[\xi_\nu]]}}{\int d\xi\ \mathrm{e}^{\beta\cos[\xi][m^{\nu}_{cc}\cos[\xi_\nu]+m^{\nu}_{sc}\sin[\xi_\nu]]+\beta\sin[\xi][m^{\nu}_{cs}\cos[\xi_\nu]+m^{\nu}_{ss}\sin[\xi_\nu]]}}\rangle_{\boldsymbol{\xi}},$$

$$m^{\mu}_{sc}= \tag{2.180}$$

$$\langle\sin[\xi_\mu]\frac{\int d\xi\cos[\xi]\mathrm{e}^{\beta\cos[\xi][m^{\nu}_{cc}\cos[\xi_\nu]+m^{\nu}_{sc}\sin[\xi_\nu]]+\beta\sin[\xi][m^{\nu}_{cs}\cos[\xi_\nu]+m^{\nu}_{ss}\sin[\xi_\nu]]}}{\int d\xi\ \mathrm{e}^{\beta\cos[\xi][m^{\nu}_{cc}\cos[\xi_\nu]+m^{\nu}_{sc}\sin[\xi_\nu]]+\beta\sin[\xi][m^{\nu}_{cs}\cos[\xi_\nu]+m^{\nu}_{ss}\sin[\xi_\nu]]}}\rangle_{\boldsymbol{\xi}},$$

$$m^{\mu}_{ss}= \tag{2.181}$$

$$\langle\sin[\xi_\mu]\frac{\int d\xi\sin[\xi]\mathrm{e}^{\beta\cos[\xi][m^{\nu}_{cc}\cos[\xi_\nu]+m^{\nu}_{sc}\sin[\xi_\nu]]+\beta\sin[\xi][m^{\nu}_{cs}\cos[\xi_\nu]+m^{\nu}_{ss}\sin[\xi_\nu]]}}{\int d\xi\ \mathrm{e}^{\beta\cos[\xi][m^{\nu}_{cc}\cos[\xi_\nu]+m^{\nu}_{sc}\sin[\xi_\nu]]+\beta\sin[\xi][m^{\nu}_{cs}\cos[\xi_\nu]+m^{\nu}_{ss}\sin[\xi_\nu]]}}\rangle_{\boldsymbol{\xi}},$$

The equilibrium values of the observables $\mathbf{m}_{\star\star}$, as defined in (2.175,2.176), are now given by the solution of the coupled equations (2.178-2.181) which minimizes

$$f(\{\mathbf{m}_{\star\star}\}) = \frac{1}{2}\sum_{\star\star}\mathbf{m}_{\star\star}^2 - \frac{1}{\beta}\langle\log\int d\xi \;\times \tag{2.182}$$
$$\mathrm{e}^{\beta\cos[\xi][m_{cc}^{\nu}\cos[\xi_\nu]+m_{sc}^{\nu}\sin[\xi_\nu]]+\beta\sin[\xi][m_{cs}^{\nu}\cos[\xi_\nu]+m_{ss}^{\nu}\sin[\xi_\nu]]}\rangle_{\boldsymbol{\xi}}.$$

We can confirm that the relevant saddle–point must be a minimum by inspecting the $\beta = 0$ limit (infinite noise levels):

$$\lim_{\beta\to 0} f(\{\mathbf{m}_{\star\star}\}) = \frac{1}{2}\sum_{\star\star}\mathbf{m}_{\star\star}^2 - \frac{1}{\beta}\log(2\pi).$$

From now on we will restrict our analysis to phase pattern components ξ_i^μ which have all been drawn independently at random from $[-\pi,\pi]$, with uniform probability density, so that $\langle g[\boldsymbol{\xi}]\rangle_{\boldsymbol{\xi}} = (2\pi)^{-p}\int_{-\pi}^{\pi}\ldots\int_{-\pi}^{\pi} d\boldsymbol{\xi}\; g[\boldsymbol{\xi}]$. At $\beta = 0$ ($T = \infty$) one finds only the trivial state $m_{\star\star}^\mu = 0$. It can be shown that there will be no discontinuous transitions to a non–trivial state as the noise level (temperature) is reduced. The continuous ones follow upon expansion of the equations (2.178-2.181) for small $\{\mathbf{m}_{\star\star}\}$, which is found to give (for each μ and each combination $\star\star$):

$$m_{\star\star}^\mu = \frac{1}{4}\beta m_{\star\star}^\mu + \mathcal{O}(\{\mathbf{m}_{\star\star}^2\}).$$

Thus a continuous transition to recall states occurs at $T = \frac{1}{4}$. Full classification of all solutions of (2.178-2.181) is ruled out. Here we will restrict ourselves to the most relevant ones, such as the pure states, where $m_{\star\star}^\mu = m_{\star\star}\delta_{\mu\lambda}$ (for some pattern label λ). Here the oscillator phases are correlated with only one of the stored phase patterns (if at all). Insertion into the above expression for $f(\{\mathbf{m}_{\star\star}\})$ shows that for such solutions we have to minimize [Coo01, SC00, SC01, CKS05]

$$f(\{m_{\star\star}\}) = \frac{1}{2}\sum_{\star\star}m_{\star\star}^2 - \frac{1}{\beta}\int\frac{d\xi}{2\pi}\;\log\int d\xi\;\times \tag{2.183}$$
$$\mathrm{e}^{\beta\cos[\xi][m_{cc}\cos[\xi]+m_{sc}\sin[\xi]]+\beta\sin[\xi][m_{cs}\cos[\xi]+m_{ss}\sin[\xi]]}.$$

We anticipate solutions corresponding to the (partial) recall of the stored phase pattern $\boldsymbol{\xi}^\lambda$ or its mirror image (modulo overall phase shifts $\xi_i \to \xi_i + \delta$, under which the synapses are obviously invariant). Insertion into (2.178-2.181) of the state

$$\xi_i = \xi_i^\lambda + \delta \quad \text{gives} \quad (m_{cc}, m_{sc}, m_{cs}, m_{ss}) = \frac{1}{2}(\cos\delta, -\sin\delta, \sin\delta, \cos\delta).$$

Similarly, insertion into (2.178-2.181) of

$$\xi_i = -\xi_i^\lambda + \delta \quad \text{gives} \quad (m_{cc}, m_{sc}, m_{cs}, m_{ss}) = \frac{1}{2}(\cos\delta, \sin\delta, \sin\delta, -\cos\delta).$$

Thus we can identify retrieval states as those solutions which are of the form

(i) retrieval of $\boldsymbol{\xi}^\lambda$: $(m_{cc}, m_{sc}, m_{cs}, m_{ss}) = m(\cos\delta, -\sin\delta, \sin\delta, \cos\delta)$,
(ii) retrieval of $-\boldsymbol{\xi}^\lambda$: $(m_{cc}, m_{sc}, m_{cs}, m_{ss}) = m(\cos\delta, \sin\delta, \sin\delta, -\cos\delta)$,

with full recall corresponding to $m = \frac{1}{2}$. Insertion into the saddle–point equations and into (2.183), followed by an appropriate shift of the integration variable ξ, shows that the free energy is independent of δ (so the above two ansätzes solve the saddle–point equations for any δ) and that

$$m = \frac{1}{2}\frac{\int d\xi\ \cos[\xi] \mathrm{e}^{\beta m \cos[\xi]}}{\int d\xi\ \mathrm{e}^{\beta m \cos[\xi]}}, \qquad f(m) = m^2 - \frac{1}{\beta}\log \int d\xi\ \mathrm{e}^{\beta m \cos[\xi]}.$$

Expansion in powers of m, using $\log(1+z) = z - \frac{1}{2}z^2 + \mathcal{O}(z^3)$, reveals that non-zero minima m indeed bifurcate continuously at $T = \beta^{-1} = \frac{1}{4}$:

$$f(m) + \frac{1}{\beta}\log[2\pi] = (1 - \frac{1}{4}\beta)m^2 + \frac{1}{64}\beta^3 m^4 + \mathcal{O}(m^6). \tag{2.184}$$

Retrieval states are obviously not the only pure states that solve the saddle–point equations. The function (2.183) is invariant under the following discrete (non-commuting) transformations:

$$\begin{aligned} &\mathrm{I}: \ (m_{cc}, m_{sc}, m_{cs}, m_{ss}) \to (m_{cc}, m_{sc}, -m_{cs}, -m_{ss}), \\ &\mathrm{II}: (m_{cc}, m_{sc}, m_{cs}, m_{ss}) \to (m_{cs}, m_{ss}, m_{cc}, m_{sc}). \end{aligned}$$

We expect these to induce solutions with specific symmetries. In particular we anticipate the following symmetric and antisymmetric states:

symmetric under I :$(m_{cc}, m_{sc}, m_{cs}, m_{ss}) = \sqrt{2}m(\cos\delta, \sin\delta, 0, 0)$,
antisymmetric under I :$(m_{cc}, m_{sc}, m_{cs}, m_{ss}) = \sqrt{2}m(0, 0, \cos\delta, \sin\delta)$,
symmetric under II :$(m_{cc}, m_{sc}, m_{cs}, m_{ss}) = m(\cos\delta, \sin\delta, \cos\delta, \sin\delta)$,
antisymmetric under II :$(m_{cc}, m_{sc}, m_{cs}, m_{ss}) = m(\cos\delta, \sin\delta, -\cos\delta, -\sin\delta)$.

Insertion into the saddle–point equations and into (2.183) shows in all four cases the parameter δ is arbitrary and that always

$$\begin{aligned} m &= \frac{1}{\sqrt{2}}\int \frac{d\xi}{2\pi}\ \cos[\xi]\ \frac{\int d\xi\ \cos[\xi] \mathrm{e}^{\beta m\sqrt{2}\cos[\xi]\cos[\xi]}}{\int d\xi\ \mathrm{e}^{\beta m\sqrt{2}\cos[\xi]\cos[\xi]}}, \\ f(m) &= m^2 - \frac{1}{\beta}\int \frac{d\xi}{2\pi}\ \log \int d\xi\ \mathrm{e}^{\beta m\sqrt{2}\cos[\xi]\cos[\xi]}. \end{aligned}$$

Expansion in powers of m reveals that non-zero solutions m here again bifurcate continuously at $T = \frac{1}{4}$:

$$f(m) + \frac{1}{\beta}\log[2\pi] = (1 - \frac{1}{4}\beta)m^2 + \frac{3}{2}\cdot\frac{1}{64}\beta^3 m^4 + \mathcal{O}(m^6). \tag{2.185}$$

However, comparison with (2.184) shows that the free energy of the pure recall states is lower. Thus the system will prefer the recall states over the above solutions with specific symmetries.

Note, that the free energy and the order parameter equation for the pure recall states can be written in terms of *Bessel functions* as follows:

$$m = \frac{1}{2}\frac{I_1(\beta m)}{I_0(\beta m)}, \qquad f(m) = m^2 - \frac{1}{\beta}\log[2\pi I_0(\beta m)]$$

2.3.6 Attractor Neural Networks with Binary Neurons

The simplest non–trivial recurrent neural networks consist of N binary neurons $\sigma_i \in \{-1, 1\}$, which respond stochastically to post-synaptic potentials (or local fields) $h_i(\boldsymbol{\sigma})$, with $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_N)$. The fields depend linearly on the instantaneous neuron states,

$$h_i(\boldsymbol{\sigma}) = J_{ij}\sigma_j + \theta_i,$$

with the J_{ij} representing synaptic efficacies, and the θ_i representing external stimuli and/or neural thresholds [Coo01, SC00, SC01, CKS05].

Closed Macroscopic Laws for Sequential Dynamics

Here, we will first show how for sequential dynamics (where neurons are updated one after the other) one can calculate, from the microscopic stochastic laws, differential equations for the probability distribution of suitably defined macroscopic observables. For mathematical convenience our starting point will be the continuous-time master equation for the microscopic probability distribution $p_t(\boldsymbol{\sigma})$

$$\begin{aligned} \dot{p}_t(\boldsymbol{\sigma}) &= \{w_i(F_i\boldsymbol{\sigma})p_t(F_i\boldsymbol{\sigma}) - w_i(\boldsymbol{\sigma})p_t(\boldsymbol{\sigma})\}\,, \\ w_i(\boldsymbol{\sigma}) &= \frac{1}{2}[1 - \sigma_i \tanh[\beta h_i(\boldsymbol{\sigma})]], \end{aligned} \tag{2.186}$$

with $F_i\Phi(\boldsymbol{\sigma}) = \Phi(\sigma_1, \ldots, \sigma_{i-1}, -\sigma_i, \sigma_{i+1}, \ldots, \sigma_N)$.

Let us illustrate the basic ideas with the help of a simple (infinite range) toy model: $J_{ij} = (J/N)\eta_i\xi_j$ and $\theta_i = 0$ (the variables η_i and ξ_i are arbitrary, but may not depend on N). For $\eta_i = \xi_i = 1$ we get a network with uniform synapses. For $\eta_i = \xi_i \in \{-1, 1\}$ and $J > 0$ we recover the Hopfield [Hop82] model with one stored pattern. Note: the synaptic matrix is non-symmetric as soon as a pair (ij) exists such that $\eta_i\xi_j \neq \eta_j\xi_i$, so in general equilibrium statistical mechanics will not apply. The local fields become $h_i(\boldsymbol{\sigma}) = J\eta_i m(\boldsymbol{\sigma})$ with $m(\boldsymbol{\sigma}) = \frac{1}{N}\sum_k \xi_k\sigma_k$. Since they depend on the microscopic state $\boldsymbol{\sigma}$ only through the value of m, the latter quantity appears to constitute a natural macroscopic level of description. The probability density of finding the macroscopic state $m(\boldsymbol{\sigma}) = m$ is given by $\mathcal{P}_t[m] = \sum_{\boldsymbol{\sigma}} p_t(\boldsymbol{\sigma})\delta[m - m(\boldsymbol{\sigma})]$. Its time derivative follows upon inserting (2.186):

$$\dot{\mathcal{P}}_t[m] = \sum_{\boldsymbol{\sigma}} p_t(\boldsymbol{\sigma}) w_k(\boldsymbol{\sigma}) \left\{ \delta[m - m(\boldsymbol{\sigma}) + \frac{2}{N}\xi_k\sigma_k] - \delta\left[m - m(\boldsymbol{\sigma})\right] \right\}$$
$$= \frac{\partial}{\partial m} \left\{ \sum_{\boldsymbol{\sigma}} p_t(\boldsymbol{\sigma}) \delta\left[m - m(\boldsymbol{\sigma})\right] \frac{2}{N}\xi_k\sigma_k w_k(\boldsymbol{\sigma}) \right\} + \mathcal{O}(\frac{1}{N}).$$

Inserting our expressions for the transition rates $w_i(\boldsymbol{\sigma})$ and the local fields $h_i(\boldsymbol{\sigma})$ gives:

$$\dot{\mathcal{P}}_t[m] = \frac{\partial}{\partial m} \left\{ \mathcal{P}_t[m] \left[m - \frac{1}{N}\xi_k \tanh[\eta_k \beta J m] \right] \right\} + \mathcal{O}(N^{-1}).$$

In the limit $N \to \infty$ only the first term survives. The general solution of the resulting Liouville equation is

$$\mathcal{P}_t[m] = \int dm_0 \; \mathcal{P}_0[m_0] \delta\left[m - m(t|m_0)\right],$$

where $m(t|m_0)$ is the solution of

$$\dot{m} = \lim_{N\to\infty} \frac{1}{N}\xi_k \tanh[\eta_k \beta J m] - m, \qquad m(0) = m_0. \tag{2.187}$$

This describes deterministic evolution; the only uncertainty in the value of m is due to uncertainty in initial conditions. If at $t = 0$ the quantity m is known exactly, this will remain the case for finite time–scales; m turns out to evolve in time according to (2.187).

Let us now allow for less trivial choices of the synaptic matrix $\{J_{ij}\}$ and try to calculate the evolution in time of a given set of macroscopic observables $\boldsymbol{\Omega}(\boldsymbol{\sigma}) = (\Omega_1(\boldsymbol{\sigma}), \ldots, \Omega_n(\boldsymbol{\sigma}))$ in the limit $N \to \infty$. There are no restrictions yet on the form or the number n of these state variables; these will, however, arise naturally if we require the observables $\boldsymbol{\Omega}$ to obey a closed set of deterministic laws, as we will see. The probability density of finding the system in macroscopic state $\boldsymbol{\Omega}$ is given by [Coo01, SC00, SC01, CKS05]:

$$\mathcal{P}_t\left[\boldsymbol{\Omega}\right] = \sum_{\boldsymbol{\sigma}} p_t(\boldsymbol{\sigma}) \delta\left[\boldsymbol{\Omega} - \boldsymbol{\Omega}(\boldsymbol{\sigma})\right]. \tag{2.188}$$

Its time derivative is got by inserting (2.186). If in those parts of the resulting expression which contain the operators F_i we perform the transformations $\boldsymbol{\sigma} \to F_i\boldsymbol{\sigma}$, we arrive at

$$\dot{\mathcal{P}}_t\left[\boldsymbol{\Omega}\right] = \sum_{\boldsymbol{\sigma}} p_t(\boldsymbol{\sigma}) w_i(\boldsymbol{\sigma}) \left\{ \delta\left[\boldsymbol{\Omega} - \boldsymbol{\Omega}(F_i\boldsymbol{\sigma})\right] - \delta\left[\boldsymbol{\Omega} - \boldsymbol{\Omega}(\boldsymbol{\sigma})\right] \right\}.$$

If we define

$$\Omega_\mu(F_i\boldsymbol{\sigma}) = \Omega_\mu(\boldsymbol{\sigma}) + \Delta_{i\mu}(\boldsymbol{\sigma})$$

and make a Taylor expansion in powers of $\{\Delta_{i\mu}(\boldsymbol{\sigma})\}$, we finally get the so–called *Kramers–Moyal expansion*:[9]

$$\dot{\mathcal{P}}_t\left[\boldsymbol{\Omega}\right] = \sum_{\ell\geq 1} \frac{(-1)^\ell}{\ell+} \sum_{\mu_1=1}^{n} \cdots \sum_{\mu_\ell=1}^{n} \frac{\partial^\ell}{\partial\Omega_{\mu_1}\cdots\partial\Omega_{\mu_\ell}} \left\{\mathcal{P}_t\left[\boldsymbol{\Omega}\right] F^{(\ell)}_{\mu_1\cdots\mu_\ell}\left[\boldsymbol{\Omega};t\right]\right\}. \tag{2.189}$$

It involves conditional averages $\langle f(\boldsymbol{\sigma})\rangle_{\boldsymbol{\Omega};t}$ and the 'discrete derivatives'

$$\begin{aligned} \Delta_{j\mu}(\boldsymbol{\sigma}) &= \Omega_\mu(F_j\boldsymbol{\sigma}) - \Omega_\mu(\boldsymbol{\sigma}), \\ F^{(l)}_{\mu_1\cdots\mu_l}\left[\boldsymbol{\Omega};t\right] &= \langle w_j(\boldsymbol{\sigma})\Delta_{j\mu_1}(\boldsymbol{\sigma})\cdots\Delta_{j\mu_\ell}(\boldsymbol{\sigma})\rangle_{\boldsymbol{\Omega};t}, \\ \langle f(\boldsymbol{\sigma})\rangle_{\boldsymbol{\Omega};t} &= \frac{\sum_{\boldsymbol{\sigma}} p_t(\boldsymbol{\sigma})\delta\left[\boldsymbol{\Omega}-\boldsymbol{\Omega}(\boldsymbol{\sigma})\right] f(\boldsymbol{\sigma})}{\sum_{\boldsymbol{\sigma}} p_t(\boldsymbol{\sigma})\delta\left[\boldsymbol{\Omega}-\boldsymbol{\Omega}(\boldsymbol{\sigma})\right]}. \end{aligned} \tag{2.190}$$

Retaining only the $\ell = 1$ term in (2.189) would lead us to a Liouville equation, which describes deterministic flow in $\boldsymbol{\Omega}$ space. Including also the $\ell = 2$ term leads us to a Fokker–Planck equation which, in addition to flow, describes diffusion of the macroscopic probability density. Thus a sufficient condition for the observables $\boldsymbol{\Omega}(\boldsymbol{\sigma})$ to evolve in time deterministically in the limit $N \to \infty$ is:

$$\lim_{N\to\infty} \sum_{\ell\geq 2} \frac{1}{\ell+} \sum_{\mu_1=1}^{n} \cdots \sum_{\mu_\ell=1}^{n} \langle|\Delta_{j\mu_1}(\boldsymbol{\sigma})\cdots\Delta_{j\mu_\ell}(\boldsymbol{\sigma})|\rangle_{\boldsymbol{\Omega};t} = 0. \tag{2.191}$$

In the simple case where all observables Ω_μ scale similarly in the sense that all 'derivatives' $\Delta_{j\mu} = \Omega_\mu(F_i\boldsymbol{\sigma}) - \Omega_\mu(\boldsymbol{\sigma})$ are of the same order in N (i.e., there is a monotonic function $\tilde{\Delta}_N$ such that $\Delta_{j\mu} = \mathcal{O}(\tilde{\Delta}_N)$ for all $j\mu$), for instance, criterion (2.191) becomes:

$$\lim_{N\to\infty} n\tilde{\Delta}_N\sqrt{N} = 0. \tag{2.192}$$

If for a given set of observables condition (2.191) is satisfied, we can for large N describe the evolution of the macroscopic probability density by a Liouville equation:

$$\dot{\mathcal{P}}_t\left[\boldsymbol{\Omega}\right] = -\frac{\partial}{\partial\Omega_\mu}\left\{\mathcal{P}_t\left[\boldsymbol{\Omega}\right] F^{(1)}_\mu\left[\boldsymbol{\Omega};t\right]\right\},$$

whose solution describes deterministic flow,

[9] The Kramers–Moyal expansion (2.189) is to be interpreted in a distributional sense, i.e., only to be used in expressions of the form $\int d\boldsymbol{\Omega}_t(\boldsymbol{\Omega})G(\boldsymbol{\Omega})$ with smooth functions $G(\boldsymbol{\Omega})$, so that all derivatives are well–defined and finite. Furthermore, (2.189) will only be useful if the $\Delta_{j\mu}$, which measure the sensitivity of the macroscopic quantities to single neuron state changes, are sufficiently small. This is to be expected: for finite N *any* observable can only assume a finite number of possible values; only for $N \to \infty$ may we expect smooth probability distributions for our macroscopic quantities.

$$\mathcal{P}_t[\mathbf{\Omega}] = \int d\mathbf{\Omega}_0 \mathcal{P}_0[\mathbf{\Omega}_0]\delta[\mathbf{\Omega} - \mathbf{\Omega}(t|\mathbf{\Omega}_0)],$$

with $\mathbf{\Omega}(t|\mathbf{\Omega}_0)$ given, in turn, as the solution of

$$\dot{\mathbf{\Omega}}(t) = \mathbf{F}^{(1)}\left[\mathbf{\Omega}(t); t\right], \qquad \mathbf{\Omega}(0) = \mathbf{\Omega}_0. \tag{2.193}$$

In taking the limit $N \to \infty$, however, we have to keep in mind that the resulting deterministic theory is got by taking this limit for *finite* t. According to (2.189) the $\ell > 1$ terms do come into play for sufficiently large times t; for $N \to \infty$, however, these times diverge by virtue of (2.191).

Now, equation (2.193) will in general not be autonomous; tracing back the origin of the explicit time dependence in the right–hand side of (2.193) one finds that to calculate $\mathbf{F}^{(1)}$ one needs to know the microscopic probability density $p_t(\boldsymbol{\sigma})$. This, in turn, requires solving equation (2.186) (which is exactly what one tries to avoid). We will now discuss a mechanism via which to eliminate the offending explicit time dependence, and to turn the observables $\mathbf{\Omega}(\boldsymbol{\sigma})$ into an autonomous level of description, governed by *closed* dynamic laws. The idea is to choose the observables $\mathbf{\Omega}(\boldsymbol{\sigma})$ in such a way that there is no explicit time dependence in the flow field $\mathbf{F}^{(1)}\left[\mathbf{\Omega}; t\right]$ (if possible). According to (2.190) this implies making sure that there exist functions $\Phi_\mu\left[\mathbf{\Omega}\right]$ such that

$$\lim_{N\to\infty} w_j(\boldsymbol{\sigma})\Delta_{j\mu}(\boldsymbol{\sigma}) = \Phi_\mu\left[\mathbf{\Omega}(\boldsymbol{\sigma})\right], \tag{2.194}$$

in which case the time dependence of $\mathbf{F}^{(1)}$ indeed drops out and the macroscopic state vector simply evolves in time according to [Coo01, SC00, SC01, CKS05]:

$$\dot{\mathbf{\Omega}} = \mathbf{\Phi}\left[\mathbf{\Omega}\right], \qquad \mathbf{\Phi} = (\Phi_1[\mathbf{\Omega}], \ldots, \Phi_n[\mathbf{\Omega}]).$$

Clearly, for this closure method to apply, a suitable separable structure of the synaptic matrix is required. If, for instance, the macroscopic observables Ω_μ depend linearly on the microscopic state variables $\boldsymbol{\sigma}$ (i.e., $\Omega_\mu(\boldsymbol{\sigma}) = \frac{1}{N}\sum_{j=1}^N \omega_{\mu j}\sigma_j$), we get with the transition rates defined in (2.186):

$$\dot{\Omega}_\mu = \lim_{N\to\infty} \frac{1}{N}\omega_{\mu j} \tanh(\beta h_j(\boldsymbol{\sigma})) - \Omega_\mu, \tag{2.195}$$

in which case the only further condition for (2.194) to hold is that all local fields $h_k(\boldsymbol{\sigma})$ must (in leading order in N) depend on the microscopic state $\boldsymbol{\sigma}$ only through the values of the observables $\mathbf{\Omega}$; since the local fields depend linearly on $\boldsymbol{\sigma}$ this, in turn, implies that the synaptic matrix must be separable: if $J_{ij} = \sum_\mu K_{i\mu}\omega_{\mu j}$ then indeed $h_i(\boldsymbol{\sigma}) = \sum_\mu K_{i\mu}\Omega_\mu(\boldsymbol{\sigma}) + \theta_i$. Next we will show how this approach can be applied to networks for which the matrix of synapses has a separable form (which includes most symmetric and non–symmetric Hebbian type attractor models). We will restrict ourselves to models with $\theta_i = 0$; introducing non–zero thresholds is straightforward and does not pose new problems.

Application to Separable Attractor Networks

We consider the following class of models, in which the interaction matrices have the form

$$J_{ij} = \frac{1}{N} Q(\boldsymbol{\xi}_i; \boldsymbol{\xi}_j), \qquad \boldsymbol{\xi}_i = (\xi_i^1, \ldots, \xi_i^p). \tag{2.196}$$

The components ξ_i^μ, representing the information ('patterns') to be stored or processed, are assumed to be drawn from a finite discrete set Λ, containing n_Λ elements (they are not allowed to depend on N). The Hopfield model [Hop82] corresponds to choosing $Q(\mathbf{x}; \mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$ and $\Lambda \equiv \{-1, 1\}$. One now introduces a partition of the system $\{1, \ldots, N\}$ into n_Λ^p so–called sublattices $I_{\boldsymbol{\eta}}$:

$$I_{\boldsymbol{\eta}} = \{i |\ \boldsymbol{\xi}_i = \boldsymbol{\eta}\}, \qquad \{1, \ldots, N\} = \bigcup_{\boldsymbol{\eta}} I_{\boldsymbol{\eta}}, \qquad (\boldsymbol{\eta} \in \Lambda^p).$$

The number of neurons in sublattice $I_{\boldsymbol{\eta}}$ is denoted by $|I_{\boldsymbol{\eta}}|$ (this number will have to be large). If we choose as our macroscopic observables the average activities ('magnetizations') within these sublattices, we are able to express the local fields h_k solely in terms of macroscopic quantities:

$$m_{\boldsymbol{\eta}}(\boldsymbol{\sigma}) = \frac{1}{|I_{\boldsymbol{\eta}}|} \sum_{i \in I_{\boldsymbol{\eta}}} \sigma_i, \qquad h_k(\boldsymbol{\sigma}) = p_{\boldsymbol{\eta}} Q\left(\boldsymbol{\xi}_k; \boldsymbol{\eta}\right) m_{\boldsymbol{\eta}}, \tag{2.197}$$

with the relative sublattice sizes $p_{\boldsymbol{\eta}} = |I_{\boldsymbol{\eta}}|/N$. If all $p_{\boldsymbol{\eta}}$ are of the same order in N (which, for example, is the case if the vectors $\boldsymbol{\xi}_i$ have been drawn at random from the set Λ^p) we may write $\Delta_{j\boldsymbol{\eta}} = \mathcal{O}(n_\Lambda^p N^{-1})$ and use (2.192). The evolution in time of the sublattice activities is then found to be deterministic in the $N \to \infty$ limit if $\lim_{N\to\infty} p/\log N = 0$. Furthermore, condition (2.194) holds, since

$$w_j(\boldsymbol{\sigma}) \Delta_{j\boldsymbol{\eta}}(\boldsymbol{\sigma}) = \tanh[\beta p_{\boldsymbol{\eta}'} Q\left(\boldsymbol{\eta}; \boldsymbol{\eta}'\right) m_{\boldsymbol{\eta}'}] - m_{\boldsymbol{\eta}}.$$

This situation is described by (2.195), and that the evolution in time of the sublattice activities is governed by the following autonomous set of differential equations [RKH88]:

$$\dot{m}_{\boldsymbol{\eta}} = \tanh[\beta p_{\boldsymbol{\eta}'} Q\left(\boldsymbol{\eta}; \boldsymbol{\eta}'\right) m_{\boldsymbol{\eta}'}] - m_{\boldsymbol{\eta}}. \tag{2.198}$$

We see that, in contrast to the equilibrium techniques as described above, here there is no need at all to require symmetry of the interaction matrix or absence of self-interactions. In the symmetric case $Q(\mathbf{x}; \mathbf{y}) = Q(\mathbf{y}; \mathbf{x})$ the system will approach equilibrium; if the kernel Q is positive definite this can be shown, for instance, by inspection of the *Lyapunov function* $\mathcal{L}\{m_{\boldsymbol{\eta}}\}$:[10]

[10] Recall that *Lyapunov function* is a function of the state variables which is bounded from below and whose value decreases monotonically during the dynamics, see e.g., [Kha92]. Its existence guarantees evolution towards a stationary state (under some weak conditions).

$$\mathcal{L}\{m_{\boldsymbol{\eta}}\} = \frac{1}{2} p_{\boldsymbol{\eta}} m_{\boldsymbol{\eta}} Q(\boldsymbol{\eta};\boldsymbol{\eta}') m_{\boldsymbol{\eta}'} p_{\boldsymbol{\eta}'} - \frac{1}{\beta} \sum_{\boldsymbol{\eta}} p_{\boldsymbol{\eta}} \log\cosh[\beta Q(\boldsymbol{\eta};\boldsymbol{\eta}') m_{\boldsymbol{\eta}'} p_{\boldsymbol{\eta}'}],$$

which is bounded from below and obeys:

$$\dot{\mathcal{L}} = -\left[p_{\boldsymbol{\eta}} \dot{m}_{\boldsymbol{\eta}}\right] Q(\boldsymbol{\eta};\boldsymbol{\eta}') \left[p_{\boldsymbol{\eta}'} \dot{m}_{\boldsymbol{\eta}'}\right] \leq 0. \tag{2.199}$$

Note that from the sublattice activities, in turn, follow the *Hopfield overlaps* $m_\mu(\boldsymbol{\sigma})$,

$$m_\mu(\boldsymbol{\sigma}) = \frac{1}{N} \xi_i^\mu \sigma_i = p_{\boldsymbol{\eta}} \eta_\mu m_{\boldsymbol{\eta}}. \tag{2.200}$$

Simple examples of relevant models of the type (2.196), the dynamics of which are for large N described by equation (2.198), are for instance the ones where one applies a nonlinear operation Φ to the standard Hopfield–type [Hop82] (or Hebbian) interactions . This nonlinearity could result from e.g., a clipping procedure or from retaining only the *sign* of the Hebbian values:

$$J_{ij} = \frac{1}{N} \Phi(\xi_i^\mu \xi_j^\mu): \qquad \text{e.g.,}$$
$$\Phi(x) = \begin{cases} -K & \text{for} \quad x \leq K \\ x & \text{for} \; -K < x < K \\ K & \text{for} \quad x \geq K \end{cases}, \qquad \text{or} \qquad \Phi(x) = \operatorname{sgn}(x).$$

The effect of introducing such nonlinearities is found to be of a quantitative nature, giving rise to little more than a re-scaling of critical noise levels and storage capacities. We will illustrate this statement by working out the $p = 2$ equations for randomly drawn pattern bits $\xi_i^\mu \in \{-1, 1\}$, where there are only four sub-lattices, and where $p_{\boldsymbol{\eta}} = \frac{1}{4}$ for all $\boldsymbol{\eta}$ (details can be found in e.g., [DHS91]). Using $\Phi(0) = 0$ and $\Phi(-x) = -\Phi(x)$ (as with the above examples) we get from (2.198):

$$\dot{m}_{\boldsymbol{\eta}} = \tanh[\frac{1}{4} \beta \Phi(2)(m_{\boldsymbol{\eta}} - m_{-\boldsymbol{\eta}})] - m_{\boldsymbol{\eta}}. \tag{2.201}$$

Here the choice made for $\Phi(x)$ shows up only as a re-scaling of the temperature. From (2.201) we further get $\frac{d}{dt}(m_{\boldsymbol{\eta}} + m_{-\boldsymbol{\eta}}) = -(m_{\boldsymbol{\eta}} + m_{-\boldsymbol{\eta}})$. The system decays exponentially towards a state where, according to (2.200), $m_{\boldsymbol{\eta}} = -m_{-\boldsymbol{\eta}}$ for all $\boldsymbol{\eta}$. If at $t = 0$ this is already the case, we find (at least for $p = 2$) decoupled equations for the sub-lattice activities.

Now, equations (2.198,2.200) suggest that at the level of overlaps there will be, in turn, closed laws if the kernel Q is bilinear, $Q(\mathbf{x};\mathbf{y}) = \sum_{\mu\nu} x_\mu A_{\mu\nu} y_\nu$, or [Coo01, SC00, SC01, CKS05]:

$$J_{ij} = \frac{1}{N} \xi_i^\mu A_{\mu\nu} \xi_j^\nu, \qquad \boldsymbol{\xi}_i = (\xi_i^1, \ldots, \xi_i^p). \tag{2.202}$$

We will see that now the ξ_i^μ need not be drawn from a finite discrete set (as long as they do not depend on N). The Hopfield model corresponds to $A_{\mu\nu} = \delta_{\mu\nu}$ and $\xi_i^\mu \in \{-1, 1\}$. The fields h_k can now be written in terms of the overlaps m_μ:

$$h_k(\boldsymbol{\sigma}) = \boldsymbol{\xi}_k \cdot A\mathbf{m}(\boldsymbol{\sigma}), \qquad \mathbf{m} = (m_1, \ldots, m_p), \qquad m_\mu(\boldsymbol{\sigma}) = \frac{1}{N}\xi_i^\mu \sigma_i.$$

For this choice of macroscopic variables we find $\Delta_{j\mu} = \mathcal{O}(N^{-1})$, so the evolution of the vector $\mathbf{m}$ becomes deterministic for $N \to \infty$ if, according to (2.192), $\lim_{N\to\infty} p/\sqrt{N} = 0$. Again (2.194) holds, since

$$w_j(\boldsymbol{\sigma})\Delta_{j\mu}(\boldsymbol{\sigma}) = \frac{1}{N}\boldsymbol{\xi}_k \tanh\left[\beta \boldsymbol{\xi}_k \cdot A\mathbf{m}\right] - \mathbf{m}.$$

Thus the evolution in time of the overlap vector $\mathbf{m}$ is governed by a closed set of differential equations:

$$\dot{\mathbf{m}} = \langle \boldsymbol{\xi} \tanh\left[\beta \boldsymbol{\xi} \cdot A\mathbf{m}\right] \rangle_{\boldsymbol{\xi}} - \mathbf{m}, \qquad \langle \Phi(\boldsymbol{\xi}) \rangle_{\boldsymbol{\xi}} = \int d\boldsymbol{\xi}\ \rho(\boldsymbol{\xi})\Phi(\boldsymbol{\xi}), \qquad (2.203)$$

with $\rho(\boldsymbol{\xi}) = \lim_{N\to\infty} N^{-1} \sum_i \delta[\boldsymbol{\xi} - \boldsymbol{\xi}_i]$. Symmetry of the synapses is not required. For certain non–symmetric matrices A one finds stable limit–cycle solutions of (2.203). In the symmetric case $A_{\mu\nu} = A_{\nu\mu}$ the system will approach equilibrium; the Lyapunov function (2.199) for positive definite matrices A now becomes:

$$\mathcal{L}\{\mathbf{m}\} = \frac{1}{2}\mathbf{m} \cdot A\mathbf{m} - \frac{1}{\beta}\langle \log\cosh\left[\beta \boldsymbol{\xi} \cdot A\mathbf{m}\right] \rangle_{\boldsymbol{\xi}}.$$

As a second simple application of the flow equations (2.203) we turn to the relaxation times corresponding to the attractors of the Hopfield model (where $A_{\mu\nu} = \delta_{\mu\nu}$). Expanding (2.203) near a stable fixed–point $\mathbf{m}^*$, i.e., $\mathbf{m}(t) = \mathbf{m}^* + \mathbf{x}(t)$ with $|\mathbf{x}(t)| \ll 1$, gives the linearized equation

$$\dot{x}_\mu = [\beta \langle \xi_\mu \xi_\nu \tanh[\beta \boldsymbol{\xi} \cdot \mathbf{m}^*] \rangle_{\boldsymbol{\xi}} - \delta_{\mu\nu}] x_\nu + \mathcal{O}(\mathbf{x}^2). \qquad (2.204)$$

The Jacobian of (2.203), which determines the linearized equation (2.204), turns out to be *minus* the curvature matrix of the free energy surface at the fixed-point. The asymptotic relaxation towards any stable attractor is generally exponential, with a characteristic time τ given by the inverse of the smallest eigenvalue of the curvature matrix. If, in particular, for the fixed–point $\mathbf{m}^*$ we substitute an $n-$mixture state, i.e., $m_\mu = m_n$ $(\mu \leq n)$ and $m_\mu = 0$ $(\mu > n)$, and transform (2.204) to the basis where the corresponding curvature matrix $\mathbf{D}^{(n)}$ (with eigenvalues D_λ^n) is diagonal, $\mathbf{x} \to \tilde{\mathbf{x}}$, we get $\tilde{x}_\lambda(t) = \tilde{x}_\lambda(0)\mathrm{e}^{-tD_\lambda^n} + \ldots$so $\tau^{-1} = \min_\lambda D_\lambda^n$, which we have already calculated in determining the character of the saddle–points of the free-energy surface. The relaxation time for the $n-$mixture attractors decreases monotonically

with the degree of mixing n, for any noise level. At the transition where a macroscopic state $\mathbf{m}^*$ ceases to correspond to a local minimum of the free energy surface, it also de–stabilizes in terms of the linearized dynamic equation (2.204) (as it should). The Jacobian develops a zero eigenvalue, the relaxation time diverges, and the long–time behavior is no longer got from the linearized equation. This gives rise to critical slowing down (i.e., power law relaxation as opposed to exponential relaxation). For instance, at the transition temperature $T_c = 1$ for the $n = 1$ (pure) state, we find by expanding (2.203):

$$\dot{m}_\mu = m_\mu[\frac{2}{3}m_\mu^2 - \mathbf{m}^2] + \mathcal{O}(\mathbf{m}^5),$$

which gives rise to a relaxation towards the trivial fixed–point of the form $\mathbf{m} \sim t^{-\frac{1}{2}}$.

If one is willing to restrict oneself to the limited class of models (2.202) (as opposed to the more general class (2.196)) and to the more global level of description in terms of p overlap parameters m_μ instead of n_Λ^p sublattice activities $m_{\boldsymbol{\eta}}$, then there are two rewards. Firstly there will be no restrictions on the stored pattern components ξ_i^μ (for instance, they are allowed to be real-valued); secondly the number p of patterns stored can be much larger for the deterministic autonomous dynamical laws to hold ($p \ll \sqrt{N}$ instead of $p \ll \log N$, which from a biological point of view is not impressive [Coo01, SC00, SC01, CKS05].

Closed Macroscopic Laws for Parallel Dynamics

We now turn to the parallel dynamics counterpart of (2.186), i.e., the Markov chain

$$p_{\ell+1}(\boldsymbol{\sigma}) = \sum_{\boldsymbol{\sigma}'} W[\boldsymbol{\sigma};\boldsymbol{\sigma}']\, p_\ell(\boldsymbol{\sigma}'), \tag{2.205}$$

$$W[\boldsymbol{\sigma};\boldsymbol{\sigma}'] = \prod_{i=1}^{N} \frac{1}{2}\left[1 + \sigma_i \tanh[\beta h_i(\boldsymbol{\sigma}')]\right],$$

with $\sigma_i \in \{-1, 1\}$, and with local fields $h_i(\boldsymbol{\sigma})$ defined in the usual way. The evolution of macroscopic probability densities will here be described by discrete maps, in stead of differential equations.

Let us first see what happens to our previous toy model: $J_{ij} = (J/N)\eta_i\xi_j$ and $\theta_i = 0$. As before we try to describe the dynamics at the (macroscopic) level of the quantity $m(\boldsymbol{\sigma}) = \frac{1}{N}\sum_k \xi_k\sigma_k$. The evolution of the macroscopic probability density $\mathcal{P}_t[m]$ is got by inserting (2.205):

$$\begin{aligned}\mathcal{P}_{t+1}[m] &= \sum_{\boldsymbol{\sigma}\boldsymbol{\sigma}'} \delta\left[m - m(\boldsymbol{\sigma})\right] W\left[\boldsymbol{\sigma};\boldsymbol{\sigma}'\right] p_t(\boldsymbol{\sigma}') \\ &= \int dm'\ \tilde{W}_t\left[m,m'\right] \mathcal{P}_t[m'], \qquad \text{with} \qquad (2.206) \\ \tilde{W}_t\left[m,m'\right] &= \frac{\sum_{\boldsymbol{\sigma}\boldsymbol{\sigma}'} \delta\left[m - m(\boldsymbol{\sigma})\right] \delta\left[m' - m(\boldsymbol{\sigma}')\right] W\left[\boldsymbol{\sigma};\boldsymbol{\sigma}'\right] p_t(\boldsymbol{\sigma}')}{\sum_{\boldsymbol{\sigma}'} \delta\left[m' - m(\boldsymbol{\sigma}')\right] p_t(\boldsymbol{\sigma}')}.\end{aligned}$$

We now insert our expression for the transition probabilities $W[\boldsymbol{\sigma};\boldsymbol{\sigma}']$ and for the local fields. Since the fields depend on the microscopic state $\boldsymbol{\sigma}$ only through $m(\boldsymbol{\sigma})$, the distribution $p_t(\boldsymbol{\sigma})$ drops out of the above expression for $\tilde{W}_t$ which thereby loses its explicit time–dependence, $\tilde{W}_t\left[m,m'\right] \to \tilde{W}\left[m,m'\right]$:

$$\begin{aligned}\tilde{W}\left[m,m'\right] &= \mathrm{e}^{-\sum_i \log\cosh(\beta J m' \eta_i)} \left\langle \delta\left[m - m(\boldsymbol{\sigma})\right] \mathrm{e}^{\beta J m' \eta_i \sigma_i} \right\rangle_{\boldsymbol{\sigma}} \\ \text{with} \qquad \left\langle \ldots \right\rangle_{\boldsymbol{\sigma}} &= 2^{-N} \sum_{\boldsymbol{\sigma}} \ldots\end{aligned}$$

Inserting the integral representation for the $\delta-$function allows us to perform the average:

$$\tilde{W}\left[m,m'\right] = \left[\frac{\beta N}{2\pi}\right] \int dk\ \mathrm{e}^{N\Psi(m,m',k)},$$

$$\Psi = i\beta k m + \left\langle \log\cosh\beta[J\eta m' - ik\xi]\right\rangle_{\eta,\xi} - \left\langle \log\cosh\beta[J\eta m']\right\rangle_{\eta}.$$

Since $\tilde{W}\left[m,m'\right]$ is (by construction) normalized, $\int dm\ \tilde{W}\left[m,m'\right] = 1$, we find that for $N \to \infty$ the expectation value with respect to $\tilde{W}\left[m,m'\right]$ of any sufficiently smooth function $f(m)$ will be determined only by the value $m^*(m')$ of m in the relevant saddle–point of Ψ:

$$\int dm\ f(m)\tilde{W}\left[m,m'\right] = \frac{\int dmdk\ f(m)\mathrm{e}^{N\Psi(m,m',k)}}{\int dmdk\ \mathrm{e}^{N\Psi(m,m',k)}} \to f(m^*(m')), \quad (N \to \infty).$$

Variation of Ψ with respect to k and m gives the two saddle–point equations:

$$m = \left\langle \xi \tanh\beta[J\eta m' - \xi k]\right\rangle_{\eta,\xi}, \qquad k = 0.$$

We may now conclude that $\lim_{N\to\infty} \tilde{W}\left[m,m'\right] = \delta\left[m - m^*(m')\right]$ with $m^*(m')$ $= \left\langle \xi \tanh(\beta J\eta m')\right\rangle_{\eta,\xi}$, and that the macroscopic equation (2.206) becomes [Coo01, SC00, SC01, CKS05]:

$$\mathcal{P}_{t+1}[m] = \int dm'\ \delta\left[m - \left\langle \xi \tanh(\beta J\eta m')\right\rangle_{\eta\xi}\right] \mathcal{P}_t[m'], \qquad (N \to \infty).$$

This describes deterministic evolution. If at $t = 0$ we know m exactly, this will remain the case for finite time-scales, and m will evolve according to a discrete version of the sequential dynamics law (2.187):

$$m_{t+1} = \left\langle \xi \tanh[\beta J\eta m_t]\right\rangle_{\eta,\xi}.$$

We now try to generalize the above approach to less trivial classes of models. As for the sequential case we will find in the limit $N \to \infty$ closed deterministic evolution equations for a more general set of intensive macroscopic state variables $\mathbf{\Omega}(\boldsymbol{\sigma}) = (\Omega_1(\boldsymbol{\sigma}), \ldots, \Omega_n(\boldsymbol{\sigma})$ if the local fields $h_i(\boldsymbol{\sigma})$ depend on the microscopic state $\boldsymbol{\sigma}$ only through the values of $\mathbf{\Omega}(\boldsymbol{\sigma})$, and if the number n of these state variables necessary to do so is not too large. The evolution of the ensemble probability density (2.188) is now got by inserting the Markov equation (2.205):

$$\mathcal{P}_{t+1}\left[\mathbf{\Omega}\right] = \int d\mathbf{\Omega}' \; \tilde{W}_t\left[\mathbf{\Omega}, \mathbf{\Omega}'\right] \mathcal{P}_t\left[\mathbf{\Omega}'\right], \tag{2.207}$$

$$\begin{aligned}\tilde{W}_t\left[\mathbf{\Omega}, \mathbf{\Omega}'\right] &= \frac{\sum_{\boldsymbol{\sigma}\boldsymbol{\sigma}'} \delta\left[\mathbf{\Omega} - \mathbf{\Omega}(\boldsymbol{\sigma})\right] \delta\left[\mathbf{\Omega}' - \mathbf{\Omega}(\boldsymbol{\sigma}')\right] W\left[\boldsymbol{\sigma}; \boldsymbol{\sigma}'\right] p_t(\boldsymbol{\sigma}')}{\sum_{\boldsymbol{\sigma}'} \delta\left[\mathbf{\Omega}' - \mathbf{\Omega}(\boldsymbol{\sigma}')\right] p_t(\boldsymbol{\sigma}')} \\ &= \left\langle \delta\left[\mathbf{\Omega} - \mathbf{\Omega}(\boldsymbol{\sigma})\right] \left\langle \mathrm{e}^{\left[\beta\sigma_i h_i(\boldsymbol{\sigma}') - \log\cosh(\beta h_i(\boldsymbol{\sigma}'))\right]} \right\rangle_{\mathbf{\Omega}';t} \right\rangle_{\boldsymbol{\sigma}},\end{aligned} \tag{2.208}$$

with $\langle \ldots \rangle_{\boldsymbol{\sigma}} = 2^{-N} \sum_{\boldsymbol{\sigma}} \ldots$, and with the conditional (or sub-shell) average defined as in (2.190). It is clear from (2.208) that in order to find autonomous macroscopic laws, i.e., for the distribution $p_t(\boldsymbol{\sigma})$ to drop out, the local fields must depend on the microscopic state $\boldsymbol{\sigma}$ only through the macroscopic quantities $\mathbf{\Omega}(\boldsymbol{\sigma})$: $h_i(\boldsymbol{\sigma}) = h_i[\mathbf{\Omega}(\boldsymbol{\sigma})]$. In this case $\tilde{W}_t$ loses its explicit time–dependence, $\tilde{W}_t\left[\mathbf{\Omega}, \mathbf{\Omega}'\right] \to \tilde{W}\left[\mathbf{\Omega}, \mathbf{\Omega}'\right]$. Inserting integral representations for the $\delta-$functions leads to:

$$\tilde{W}\left[\mathbf{\Omega}, \mathbf{\Omega}'\right] = \left[\frac{\beta N}{2\pi}\right]^n \int d\mathbf{K} \; \mathrm{e}^{N\Psi(\mathbf{\Omega}, \mathbf{\Omega}', \mathbf{K})},$$

$$\begin{aligned}\Psi = i\beta\mathbf{K}\cdot\mathbf{\Omega} + \frac{1}{N}\log\left\langle \mathrm{e}^{\beta\left[\sum_i \sigma_i h_i[\mathbf{\Omega}'] - iN\mathbf{K}\cdot\mathbf{\Omega}(\boldsymbol{\sigma})\right]}\right\rangle_{\boldsymbol{\sigma}} \\ -\frac{1}{N}\sum_i \log\cosh[\beta h_i[\mathbf{\Omega}']].\end{aligned}$$

Using the normalization $\int d\mathbf{\Omega} \; \tilde{W}\left[\mathbf{\Omega}, \mathbf{\Omega}'\right] = 1$, we can write expectation values with respect to $\tilde{W}\left[\mathbf{\Omega}, \mathbf{\Omega}'\right]$ of macroscopic quantities $f[\mathbf{\Omega}]$ as

$$\int d\mathbf{\Omega} \; f[\mathbf{\Omega}]\tilde{W}\left[\mathbf{\Omega}, \mathbf{\Omega}'\right] = \frac{\int d\mathbf{\Omega} d\mathbf{K} \; f[\mathbf{\Omega}]\mathrm{e}^{N\Psi(\mathbf{\Omega}, \mathbf{\Omega}', \mathbf{K})}}{\int d\mathbf{\Omega} d\mathbf{K} \; \mathrm{e}^{N\Psi(\mathbf{\Omega}, \mathbf{\Omega}', \mathbf{K})}}. \tag{2.209}$$

For saddle–point arguments to apply in determining the leading order in N of (2.209), we encounter restrictions on the number n of our macroscopic quantities (as expected), since n determines the dimension of the integrations in (2.209). The restrictions can be found by expanding Ψ around its maximum Ψ^*. After defining $\mathbf{x} = (\mathbf{\Omega}, \mathbf{K})$, of dimension $2n$, and after translating the location of the maximum to the origin, one has [Coo01, SC00, SC01, CKS05]

$$\Psi(\mathbf{x}) = \Psi^* - \frac{1}{2}x_\mu x_\nu H_{\mu\nu} + x_\mu x_\nu x_\rho L_{\mu\nu\rho} + \mathcal{O}(\mathbf{x}^4), \qquad \text{giving}$$

$$\frac{\int d\mathbf{x}\ g(\mathbf{x})\mathrm{e}^{N\Psi(\mathbf{x})}}{\int d\mathbf{x}\ g(\mathbf{x})\mathrm{e}^{N\Psi(\mathbf{x})}} - g(\mathbf{0}) = \frac{\int d\mathbf{x}\ [g(\mathbf{x}) - g(\mathbf{0})]\mathrm{e}^{-\frac{1}{2}N\mathbf{x}\cdot\mathbf{H}\mathbf{x}+Nx_\mu x_\nu x_\rho L_{\mu\nu\rho}+\mathcal{O}(N\mathbf{x}^4)}}{\int d\mathbf{x}\ \mathrm{e}^{-\frac{1}{2}N\mathbf{x}\cdot\mathbf{H}\mathbf{x}+Nx_\mu x_\nu x_\rho L_{\mu\nu\rho}+\mathcal{O}(N\mathbf{x}^4)}}$$

$$= \frac{\int d\mathbf{y}\ [g(\mathbf{y}/\sqrt{N}) - g(\mathbf{0})]\mathrm{e}^{-\frac{1}{2}\mathbf{y}\cdot\mathbf{H}\mathbf{y}+y_\mu y_\nu y_\rho L_{\mu\nu\rho}/\sqrt{N}+\mathcal{O}(\mathbf{y}^4/N)}}{\int d\mathbf{y}\ \mathrm{e}^{-\frac{1}{2}\mathbf{y}\cdot\mathbf{H}\mathbf{y}+y_\mu y_\nu y_\rho L_{\mu\nu\rho}/\sqrt{N}+\mathcal{O}(\mathbf{y}^4/N)}} =$$

$$\frac{\int d\mathbf{y}\ \left[N^{-\frac{1}{2}}\mathbf{y}\cdot\boldsymbol{\nabla}g(\mathbf{0}) + \mathcal{O}(\mathbf{y}^2/N)\right]\mathrm{e}^{-\frac{1}{2}\mathbf{y}\cdot\mathbf{H}\mathbf{y}}\left[1 + y_\mu y_\nu y_\rho L_{\mu\nu\rho}/\sqrt{N} + \mathcal{O}(\mathbf{y}^6/N)\right]}{\int d\mathbf{y}\ \mathrm{e}^{-\frac{1}{2}\mathbf{y}\cdot\mathbf{H}\mathbf{y}}\left[1 + y_\mu y_\nu y_\rho L_{\mu\nu\rho}/\sqrt{N} + \mathcal{O}(\mathbf{y}^6/N)\right]}$$

$$= \mathcal{O}(n^2/N) + \mathcal{O}(n^4/N^2) + \ \text{non}-\text{dominant terms}, \qquad (N, n \to \infty),$$

with $\mathbf{H}$ denoting the Hessian (curvature) matrix of the surface Ψ at the minimum Ψ^*. We thus find

$$\lim_{N\to\infty} n/\sqrt{N} = 0: \qquad \lim_{N\to\infty}\int d\mathbf{\Omega}\ f[\mathbf{\Omega}]\tilde{W}\left[\mathbf{\Omega}, \mathbf{\Omega}'\right] = f\left[\mathbf{\Omega}^*(\mathbf{\Omega}')\right],$$

where $\mathbf{\Omega}^*(\mathbf{\Omega}')$ denotes the value of $\mathbf{\Omega}$ in the saddle–point where Ψ is minimized. Variation of Ψ with respect to $\mathbf{\Omega}$ and $\mathbf{K}$ gives the saddle–point equations:

$$\mathbf{\Omega} = \frac{\langle\mathbf{\Omega}(\boldsymbol{\sigma})\mathrm{e}^{\beta\left[\sigma_i h_i[\mathbf{\Omega}'] - iN\mathbf{K}\cdot\mathbf{\Omega}(\boldsymbol{\sigma})\right]}\rangle_{\boldsymbol{\sigma}}}{\langle\mathrm{e}^{\beta[\sigma_i h_i[\mathbf{\Omega}'] - iN\mathbf{K}\cdot\mathbf{\Omega}(\boldsymbol{\sigma})]}\rangle_{\boldsymbol{\sigma}}}, \qquad \mathbf{K} = 0.$$

We may now conclude that $\lim_{N\to\infty}\tilde{W}\left[\mathbf{\Omega}, \mathbf{\Omega}'\right] = \delta\left[\mathbf{\Omega} - \mathbf{\Omega}^*(\mathbf{\Omega}')\right]$, with

$$\mathbf{\Omega}^*(\mathbf{\Omega}') = \frac{\langle\mathbf{\Omega}(\boldsymbol{\sigma})\mathrm{e}^{\beta\sigma_i h_i[\mathbf{\Omega}']}\rangle_{\boldsymbol{\sigma}}}{\langle\mathrm{e}^{\beta\sigma_i h_i[\mathbf{\Omega}']}\rangle_{\boldsymbol{\sigma}}},$$

and that for $N \to \infty$ the macroscopic equation (2.207) becomes $\mathcal{P}_{t+1}[\mathbf{\Omega}] = \int d\mathbf{\Omega}'\ \delta[\mathbf{\Omega} - \mathbf{\Omega}^*(\mathbf{\Omega}')]\mathcal{P}_t[\mathbf{\Omega}']$. This relation again describes deterministic evolution. If at $t = 0$ we know $\mathbf{\Omega}$ exactly, this will remain the case for finite time–scales and $\mathbf{\Omega}$ will evolve according to

$$\mathbf{\Omega}(t+1) = \frac{\langle\mathbf{\Omega}(\boldsymbol{\sigma})\mathrm{e}^{\beta\sigma_i h_i[\mathbf{\Omega}(t)]}\rangle_{\boldsymbol{\sigma}}}{\langle\mathrm{e}^{\beta\sigma_i h_i[\mathbf{\Omega}(t)]}\rangle_{\boldsymbol{\sigma}}}. \tag{2.210}$$

As with the sequential case, in taking the limit $N \to \infty$ we have to keep in mind that the resulting laws apply to finite t, and that for sufficiently large times terms of higher order in N do come into play. As for the sequential case, a more rigorous and tedious analysis shows that the restriction $n/\sqrt{N} \to 0$ can in fact be weakened to $n/N \to 0$. Finally, for macroscopic quantities $\mathbf{\Omega}(\boldsymbol{\sigma})$ which are linear in $\boldsymbol{\sigma}$, the remaining $\boldsymbol{\sigma}-$averages become trivial, so that [Ber91]:

$$\Omega_\mu(\boldsymbol{\sigma}) = \frac{1}{N}\omega_{\mu i}\sigma_i: \qquad \Omega_\mu(t+1) = \lim_{N\to\infty}\frac{1}{N}\omega_{\mu i}\tanh\left[\beta h_i[\mathbf{\Omega}(t)]\right].$$

Application to Separable Attractor Networks

The separable attractor models (2.196), described at the level of sublattice activities (2.197), indeed have the property that all local fields can be written in terms of the macroscopic observables. What remains to ensure deterministic evolution is meeting the condition on the number of sublattices. If all relative sublattice sizes $p_{\boldsymbol{\eta}}$ are of the same order in N (as for randomly drawn patterns) this condition again translates into $\lim_{N\to\infty} p/\log N = 0$ (as for sequential dynamics). Since the sublattice activities are linear functions of the σ_i, their evolution in time is governed by equation (2.210), which acquires the form:

$$m_{\boldsymbol{\eta}}(t+1) = \tanh[\beta p_{\boldsymbol{\eta}'} Q\left(\boldsymbol{\eta};\boldsymbol{\eta}'\right) m_{\boldsymbol{\eta}'}(t)]. \tag{2.211}$$

As for sequential dynamics, symmetry of the interaction matrix does not play a role.

At the more global level of overlaps $m_\mu(\boldsymbol{\sigma}) = N^{-1}\sum_i \xi_i^\mu \sigma_i$ we, in turn, get autonomous deterministic laws if the local fields $h_i(\boldsymbol{\sigma})$ can be expressed in terms if $\mathbf{m}(\boldsymbol{\sigma})$ only, as for the models (2.202) (or, more generally, for all models in which the interactions are of the form $J_{ij} = \sum_{\mu\leq p} f_{i\mu}\xi_j^\mu$), and with the following restriction on the number p of embedded patterns: $\lim_{N\to\infty} p/\sqrt{N} = 0$ (as with sequential dynamics). For the bilinear models (2.202), the evolution in time of the overlap vector $\mathbf{m}$ (which depends linearly on the σ_i) is governed by (2.210), which now translates into the iterative map:

$$\mathbf{m}(t+1) = \langle \boldsymbol{\xi} \tanh[\beta \boldsymbol{\xi}\cdot A\mathbf{m}(t)]\rangle_{\boldsymbol{\xi}}, \tag{2.212}$$

with $\rho(\boldsymbol{\xi})$ as defined in (2.203). Again symmetry of the synapses is not required. For parallel dynamics it is far more difficult than for sequential dynamics to construct Lyapunov functions, and prove that the macroscopic laws (2.212) for symmetric systems evolve towards a stable fixed–point (as one would expect), but it can still be done. For non–symmetric systems the macroscopic laws (2.212) can in principle display all the interesting, but complicated, phenomena of non–conservative nonlinear systems. Nevertheless, it is also not uncommon that the equations (2.212) for non–symmetric systems can be mapped by a time–dependent transformation onto the equations for related symmetric systems (mostly variants of the original Hopfield model).

Note that the fixed–points of the macroscopic equations (2.198) and (2.203) (derived for sequential dynamics) are identical to those of (2.211) and (2.212) (derived for parallel dynamics). The stability properties of these fixed–points, however, need not be the same, and have to be assessed on a case–by–case basis. For the Hopfield model, i.e., equations (2.203,2.212) with $A_{\mu\nu} = \delta_{\mu\nu}$, they are found to be the same, but already for $A_{\mu\nu} = -\delta_{\mu\nu}$ the two types of dynamics would behave differently [Coo01, SC00, SC01, CKS05].

2.3.7 Attractor Neural Networks with Continuous Neurons

Closed Macroscopic Laws

We have seen above that models of recurrent neural networks with continuous neural variables (e.g., graded–response neurons or coupled oscillators) can often be described by a Fokker–Planck equation for the microscopic state probability density $p_t(\boldsymbol{\sigma})$:

$$\dot{p}_t(\boldsymbol{\sigma}) = -\frac{\partial}{\partial \sigma_i}\left[p_t(\boldsymbol{\sigma}) f_i(\boldsymbol{\sigma})\right] + T\sum_i \frac{\partial^2}{\partial \sigma_i^2} p_t(\boldsymbol{\sigma}).$$

Averages over $p_t(\boldsymbol{\sigma})$ are denoted by $\langle G\rangle = \int d\boldsymbol{\sigma}\; p_t(\boldsymbol{\sigma}) G(\boldsymbol{\sigma}, t)$. From (2.131) one gets directly (through integration by parts) an equation for the time derivative of averages [Coo01, SC00, SC01, CKS05]:

$$\frac{d}{dt}\langle G\rangle = \langle \frac{\partial G}{\partial t}\rangle + \langle \left[f_i(\boldsymbol{\sigma}) + T\frac{\partial}{\partial \sigma_i}\right]\frac{\partial G}{\partial \sigma_i}\rangle. \tag{2.213}$$

In particular, if we apply (2.213) to $G(\boldsymbol{\sigma}, t) = \delta[\boldsymbol{\Omega} - \boldsymbol{\Omega}(\boldsymbol{\sigma})]$, for any set of macroscopic observables $\boldsymbol{\Omega}(\boldsymbol{\sigma}) = (\Omega_1(\boldsymbol{\sigma}), \ldots, \Omega_n(\boldsymbol{\sigma}))$ (in the spirit of the previous section), we get a dynamic equation for the macroscopic probability density $P_t(\boldsymbol{\Omega}) = \langle \delta[\boldsymbol{\Omega} - \boldsymbol{\Omega}(\boldsymbol{\sigma})]\rangle$, which is again of the Fokker–Planck form:

$$\begin{aligned}\dot{P}_t(\boldsymbol{\Omega}) = &-\frac{\partial}{\partial \Omega_\mu}\left\{P_t(\boldsymbol{\Omega})\;\langle\left[f_i(\boldsymbol{\sigma}) + T\frac{\partial}{\partial \sigma_i}\right]\frac{\partial}{\partial \sigma_i}\Omega_\mu(\boldsymbol{\sigma})\rangle_{\boldsymbol{\Omega};t}\right\} \\ &+ T\frac{\partial^2}{\partial \Omega_\mu \partial \Omega_\nu}\left\{P_t(\boldsymbol{\Omega})\;\langle\left[\frac{\partial}{\partial \sigma_i}\Omega_\mu(\boldsymbol{\sigma})\right]\left[\frac{\partial}{\partial \sigma_i}\Omega_\nu(\boldsymbol{\sigma})\right]\rangle_{\boldsymbol{\Omega};t}\right\},\end{aligned} \tag{2.214}$$

with the conditional (or sub–shell) averages:

$$\langle G(\boldsymbol{\sigma})\rangle_{\boldsymbol{\Omega},t} = \frac{\int d\boldsymbol{\sigma}\; p_t(\boldsymbol{\sigma})\delta[\boldsymbol{\Omega} - \boldsymbol{\Omega}(\boldsymbol{\sigma})]G(\boldsymbol{\sigma})}{\int d\boldsymbol{\sigma}\; p_t(\boldsymbol{\sigma})\delta[\boldsymbol{\Omega} - \boldsymbol{\Omega}(\boldsymbol{\sigma})]}.$$

From (2.214) we infer that a sufficient condition for the observables $\boldsymbol{\Omega}(\boldsymbol{\sigma})$ to evolve in time deterministically (i.e., for having vanishing diffusion matrix elements in (2.214)) in the limit $N \to \infty$ is

$$\lim_{N\to\infty}\langle \sum_i \left[\sum_\mu \left|\frac{\partial}{\partial \sigma_i}\Omega_\mu(\boldsymbol{\sigma})\right|\right]^2 \rangle_{\boldsymbol{\Omega};t} = 0. \tag{2.215}$$

If (2.215) holds, the macroscopic Fokker–Planck equation (2.214) reduces for $N \to \infty$ to a Liouville equation, and the observables $\boldsymbol{\Omega}(\boldsymbol{\sigma})$ will evolve in time according to the coupled deterministic equations:

$$\dot{\Omega}_\mu = \lim_{N\to\infty}\langle\left[f_i(\boldsymbol{\sigma}) + T\frac{\partial}{\partial \sigma_i}\right]\frac{\partial}{\partial \sigma_i}\Omega_\mu(\boldsymbol{\sigma})\rangle_{\boldsymbol{\Omega};t}. \tag{2.216}$$

The deterministic macroscopic equation (2.216), together with its associated condition for validity (2.215) will form the basis for the subsequent analysis.

The general derivation given above went smoothly. However, the equations (2.216) are not yet closed. It turns out that to achieve closure even for simple continuous networks we can no longer get away with just a finite (small) number of macroscopic observables (as with binary neurons). This we will now illustrate with a simple toy network of graded–response neurons:

$$\dot{u}_i(t) = J_{ij}\ g[u_j(t)] - u_i(t) + \eta_i(t),$$

with $g[z] = \frac{1}{2}[\tanh(\gamma z) + 1]$ and with the standard Gaussian white noise $\eta_i(t)$. In the language of (2.131) this means

$$f_i(\mathbf{u}) = J_{ij} g[u_j] - u_i.$$

We choose uniform synapses $J_{ij} = J/N$, so $f_i(\mathbf{u}) \to (J/N)\sum_j g[u_j] - u_i$. If (2.215) were to hold, we would find the deterministic macroscopic laws

$$\dot{\Omega}_\mu = \lim_{N\to\infty} \langle \sum_i [\frac{J}{N}\sum_j g[u_j] - u_i + T\frac{\partial}{\partial u_i}]\frac{\partial}{\partial u_i}\Omega_\mu(\mathbf{u})\rangle_{\mathbf{\Omega};t}. \tag{2.217}$$

In contrast to similar models with binary neurons, choosing as our macroscopic level of description $\mathbf{\Omega}(\mathbf{u})$ again simply the average $m(\mathbf{u}) = N^{-1}\sum_i u_i$ now leads to an equation which fails to close [Coo01, SC00, SC01, CKS05]:

$$\dot{m} = \lim_{N\to\infty} J\ \langle \frac{1}{N}\sum_j g[u_j]\rangle_{m;t} - m.$$

The term $N^{-1}\sum_j g[u_j]$ cannot be written as a function of $N^{-1}\sum_i u_i$. We might be tempted to try dealing with this problem by just including the offending term in our macroscopic set, and choose $\mathbf{\Omega}(\mathbf{u}) = (N^{-1}\sum_i u_i, N^{-1}\sum_i g[u_i])$. This would indeed solve our closure problem for the $m-$equation, but we would now find a new closure problem in the equation for the newly introduced observable. The only way out is to choose an observable *function*, namely the distribution of potentials

$$\rho(u;\mathbf{u}) = \frac{1}{N}\sum_i \delta[u - u_i], \tag{2.218}$$

$$\rho(u) = \langle \rho(u;\mathbf{u})\rangle = \langle \frac{1}{N}\sum_i \delta[u - u_i]\rangle.$$

This is to be done with care, in view of our restriction on the number of observables: we evaluate (2.218) at first only for n specific values u_μ and take the limit $n \to \infty$ only after the limit $N \to \infty$. Thus we define $\Omega_\mu(\mathbf{u}) = \frac{1}{N}\sum_i \delta[u_\mu - u_i]$, condition (2.215) reduces to the familiar expression $\lim_{N\to\infty} n/\sqrt{N} = 0$, and we get for $N \to \infty$ and $n \to \infty$ (taken in that

order) from (2.217) a diffusion equation for the distribution of membrane potentials (describing a so–called '*time–dependent Ornstein–Uhlenbeck process*' [Kam92, Gar85]):

$$\dot{\rho}(u) = -\frac{\partial}{\partial u}\left\{\rho(u)\left[J\int du'\ \rho(u')g[u'] - u\right]\right\} + T\frac{\partial^2}{\partial u^2}\rho(u). \qquad (2.219)$$

The natural solution of (2.219) [11] is the Gaussian distribution

$$\rho_t(u) = [2\pi\Sigma^2(t)]^{-\frac{1}{2}}\mathrm{e}^{-\frac{1}{2}[u-\overline{u}(t)]^2/\Sigma^2(t)}, \qquad (2.220)$$

in which $\Sigma = [T + (\Sigma_0^2 - T)\mathrm{e}^{-2t}]^{\frac{1}{2}}$, and $\overline{u}$ evolves in time according to

$$\frac{d}{dt}\overline{u} = J\int Dz\ g[\overline{u} + \Sigma z] - \overline{u},$$

with $Dz = (2\pi)^{-\frac{1}{2}}\mathrm{e}^{-\frac{1}{2}z^2}dz$. We can now also calculate the distribution $p(s)$ of neuronal firing activities $s_i = g[u_i]$ at any time,

$$p(s) = \int du\ \rho(u)\ \delta[s - g[u]] = \frac{\rho(g^{\mathrm{inv}}[s])}{\int_0^1 ds'\ \rho(g^{\mathrm{inv}}[s'])}.$$

For our choice $g[z] = \frac{1}{2} + \frac{1}{2}\tanh[\gamma z]$ we have $g^{\mathrm{inv}}[s] = \frac{1}{2\gamma}\log[s/(1-s)]$, so in combination with (2.220)[12]

$$0 < s < 1: \qquad p(s) = \frac{\mathrm{e}^{-\frac{1}{2}[(2\gamma)^{-1}\log[s/(1-s)]-\overline{u}]^2/\Sigma^2}}{\int_0^1 ds'\ \mathrm{e}^{-\frac{1}{2}[(2\gamma)^{-1}\log[s'/(1-s')]-\overline{u}]^2/\Sigma^2}}.$$

Application to Graded–Response Attractor Networks

We will now turn to attractor networks with graded–response neurons of the type (2.132), in which p binary patterns $\boldsymbol{\xi}^\mu = (\xi_1^\mu, \ldots, \xi_N^\mu) \in \{-1, 1\}^N$ have been stored via separable Hebbian synapses (2.202): $J_{ij} = (2/N)\xi_i^\mu A_{\mu\nu}\xi_j^\nu$ (the extra factor 2 is inserted for future convenience). Adding suitable thresholds $\theta_i = -\frac{1}{2}\sum_j J_{ij}$ to the right–hand sides of (2.132), and choosing the nonlinearity $g(z) = \frac{1}{2}(1 + \tanh[\gamma z])$ would then give us [Coo01, SC00, SC01, CKS05]

[11] For non-Gaussian initial conditions $\rho_0(u)$ the solution of (2.219) would in time converge towards the Gaussian solution.

[12] None of the above results (not even those on the stationary state) could have been got within equilibrium statistical mechanics, since any network of connected graded–response neurons will violate detailed balance. Secondly, there appears to be a qualitative difference between simple networks (e.g., $J_{ij} = J/N$) of binary neurons versus those of continuous neurons, in terms of the types of macroscopic observables needed for deriving closed deterministic laws: a single number $m = N^{-1}\sum_i \sigma_i$ versus a distribution $\rho(\sigma) = N^{-1}\sum_i \delta[\sigma - \sigma_i]$.

$$\dot{u}_i(t) = \sum_{\mu\nu} \xi_i^\mu A_{\mu\nu} \frac{1}{N} \sum_j \xi_j^\nu \tanh[\gamma u_j(t)] - u_i(t) + \eta_i(t),$$

so the deterministic forces are $f_i(\mathbf{u}) = N^{-1} \sum_{\mu\nu} \xi_i^\mu A_{\mu\nu} \sum_j \xi_j^\nu \tanh[\gamma u_j] - u_i$. Choosing our macroscopic observables $\mathbf{\Omega}(\mathbf{u})$ such that (2.215) holds, would lead to the deterministic macroscopic laws

$$\dot{\Omega}_\mu = \lim_{N\to\infty} \sum_{\mu\nu} A_{\mu\nu} \langle \left[\frac{1}{N} \xi_j^\nu \tanh[\gamma u_j] \right] \left[\xi_i^\mu \frac{\partial}{\partial u_i} \Omega_\mu(\mathbf{u}) \right] \rangle_{\mathbf{\Omega};t} \qquad (2.221)$$
$$+ \lim_{N\to\infty} \langle \left[T \frac{\partial}{\partial u_i} - u_i \right] \frac{\partial}{\partial u_i} \Omega_\mu(\mathbf{u}) \rangle_{\mathbf{\Omega};t}.$$

As with the uniform synapses case, the main problem to be dealt with is how to choose the $\Omega_\mu(\mathbf{u})$ such that (2.221) closes. It turns out that the canonical choice is to turn to the distributions of membrane potentials within each of the 2^p sub–lattices, as introduced above :

$$I_{\boldsymbol{\eta}} = \{i |\ \boldsymbol{\xi}_i = \boldsymbol{\eta}\} : \qquad \rho_{\boldsymbol{\eta}}(u;\mathbf{u}) = \frac{1}{|I_{\boldsymbol{\eta}}|} \sum_{i\in I_{\boldsymbol{\eta}}} \delta[u - u_i], \qquad (2.222)$$
$$\rho_{\boldsymbol{\eta}}(u) = \langle \rho_{\boldsymbol{\eta}}(u;\mathbf{u}) \rangle,$$

with $\boldsymbol{\eta} \in \{-1,1\}^p$ and $\lim_{N\to\infty} |I_{\boldsymbol{\eta}}|/N = p_{\boldsymbol{\eta}}$. Again we evaluate the distributions in (2.222) at first only for n specific values u_μ and send $n \to \infty$ after $N \to \infty$. Now condition (2.215) reduces to $\lim_{N\to\infty} 2^p/\sqrt{N} = 0$. We will keep p finite, for simplicity. Using identities such as $\sum_i \ldots = \sum_{\boldsymbol{\eta}} \sum_{i\in I_{\boldsymbol{\eta}}} \ldots$ and $i \in I_{\boldsymbol{\eta}}$:

$$\frac{\partial}{\partial u_i} \rho_{\boldsymbol{\eta}}(u;\mathbf{u}) = -|I_{\boldsymbol{\eta}}|^{-1} \frac{\partial}{\partial u} \delta[u - u_i], \qquad \frac{\partial^2}{\partial u_i^2} \rho_{\boldsymbol{\eta}}(u;\mathbf{u}) = |I_{\boldsymbol{\eta}}|^{-1} \frac{\partial^2}{\partial u^2} \delta[u - u_i],$$

we then get for $N \to \infty$ and $n \to \infty$ (taken in that order) from equation (2.221) 2^p coupled diffusion equations for the distributions $\rho_{\boldsymbol{\eta}}(u)$ of membrane potentials in each of the 2^p sub-lattices $I_{\boldsymbol{\eta}}$:

$$\dot{\rho}_{\boldsymbol{\eta}}(u) = -\frac{\partial}{\partial u} \{ \rho_{\boldsymbol{\eta}}(u) [\eta_\mu A_{\mu\nu} p_{\boldsymbol{\eta}'} \eta'_\nu \int du'\ \rho_{\boldsymbol{\eta}'}(u') \tanh[\gamma u'] - u] \} + T \frac{\partial^2}{\partial u^2} \rho_{\boldsymbol{\eta}}(u). \qquad (2.223)$$

Equation (2.223) is the basis for our further analysis. It can be simplified only if we make additional assumptions on the system's initial conditions, such as $\delta-$distributed or Gaussian distributed $\rho_{\boldsymbol{\eta}}(u)$ at $t = 0$ (see below); otherwise it will have to be solved numerically.

It is clear that (2.223) is again of the time–dependent Ornstein–Uhlenbeck form, and will thus again have Gaussian solutions as the natural ones [Coo01, SC00, SC01, CKS05]:

$$\rho_{t,\boldsymbol{\eta}}(u) = [2\pi \Sigma_{\boldsymbol{\eta}}^2(t)]^{-\frac{1}{2}} \mathrm{e}^{-\frac{1}{2}[u - \overline{u}_{\boldsymbol{\eta}}(t)]^2 / \Sigma_{\boldsymbol{\eta}}^2(t)},$$

in which $\Sigma_{\boldsymbol{\eta}}(t) = [T + (\Sigma^2_{\boldsymbol{\eta}}(0) - T)\mathrm{e}^{-2t}]^{\frac{1}{2}}$, and with the $\overline{u}_{\boldsymbol{\eta}}(t)$ evolving in time according to

$$\frac{d}{dt}\overline{u}_{\boldsymbol{\eta}} = p_{\boldsymbol{\eta}'}(\boldsymbol{\eta}\cdot\mathbf{A}\boldsymbol{\eta}')\int Dz\ \tanh[\gamma(\overline{u}_{\boldsymbol{\eta}'} + \Sigma_{\boldsymbol{\eta}'}z)] - \overline{u}_{\boldsymbol{\eta}}. \tag{2.224}$$

Our problem has thus been reduced successfully to the study of the 2^p coupled scalar equations (2.224). We can also measure the correlation between the firing activities $s_i(u_i) = \frac{1}{2}[1 + \tanh(\gamma u_i)]$ and the pattern components (similar to the overlaps in the case of binary neurons). If the pattern bits are drawn at random, i.e., $\lim_{N\to\infty}|I_{\boldsymbol{\eta}}|/N = p_{\boldsymbol{\eta}} = 2^{-p}$ for all $\boldsymbol{\eta}$, we can define a 'graded–response' equivalent $m_\mu(\mathbf{u}) = 2N^{-1}\xi^\mu_i s_i(u_i) \in [-1,1]$ of the Hopfield pattern overlaps:

$$\begin{aligned} m_\mu(\mathbf{u}) &= \frac{2}{N}\xi^\mu_i s_i(\mathbf{u}) = \frac{1}{N}\xi^\mu_i\tanh(\gamma u_i) + \mathcal{O}(N^{-\frac{1}{2}}) \\ &= p_{\boldsymbol{\eta}}\ \eta_\mu\int du\ \rho_{\boldsymbol{\eta}}(u;\mathbf{u})\tanh(\gamma u) + \mathcal{O}(N^{-\frac{1}{2}}). \end{aligned}$$

Full recall of pattern μ implies $s_i(u_i) = \frac{1}{2}[\xi^\mu_i + 1]$, giving $m_\mu(\mathbf{u}) = 1$. Since the distributions $\rho_{\boldsymbol{\eta}}(u)$ obey deterministic laws for $N\to\infty$, the same will be true for the overlaps $\mathbf{m} = (m_1,\ldots,m_p)$. For the Gaussian solutions (2.224) of (2.223) we can now proceed to replace the 2^p macroscopic laws (2.224), which reduce to $\frac{d}{dt}\overline{u}_{\boldsymbol{\eta}} = \boldsymbol{\eta}\cdot\mathbf{A}\mathbf{m} - \overline{u}_{\boldsymbol{\eta}}$ and give $\overline{u}_{\boldsymbol{\eta}} = \overline{u}_{\boldsymbol{\eta}}(0)\mathrm{e}^{-t} + \boldsymbol{\eta}\cdot\mathbf{A}\int_0^t ds\ \mathrm{e}^{s-t}\mathbf{m}(s)$, by p integral equations in terms of overlaps only:

$$\begin{aligned} m_\mu(t) = p_{\boldsymbol{\eta}}\ \eta_\mu\int Dz\ \tanh[\gamma(\overline{u}_{\boldsymbol{\eta}}(0)\mathrm{e}^{-t} \\ + \boldsymbol{\eta}\cdot\mathbf{A}\int_0^t ds\ \mathrm{e}^{s-t}\mathbf{m}(s) + z\sqrt{T + (\Sigma^2_{\boldsymbol{\eta}}(0) - T)\mathrm{e}^{-2t}})], \end{aligned} \tag{2.225}$$

with $Dz = (2\pi)^{-\frac{1}{2}}\mathrm{e}^{-\frac{1}{2}z^2}dz$. Here the sub–lattices only come in via the initial conditions.

The equations describing the asymptotic (stationary) state can be written entirely without sub-lattices,[13]

$$\begin{aligned} m_\mu &= \langle\xi_\mu\int Dz\ \tanh[\gamma(\boldsymbol{\xi}\cdot\mathbf{A}\mathbf{m} + z\sqrt{T})]\rangle_{\boldsymbol{\xi}}, \\ \rho_{\boldsymbol{\eta}}(u) &= [2\pi T]^{-\frac{1}{2}}\mathrm{e}^{-\frac{1}{2}[u-\boldsymbol{\eta}\cdot\mathbf{A}\mathbf{m}]^2/T}, \end{aligned} \tag{2.226}$$

by taking the $t\to\infty$ limit in (2.225), using $\overline{u}_{\boldsymbol{\eta}}\to\boldsymbol{\eta}\cdot\mathbf{A}\mathbf{m}$, $\Sigma_{\boldsymbol{\eta}}\to\sqrt{T}$, and the familiar notation

[13] Note the appealing similarity with previous results on networks with binary neurons in equilibrium. For $T = 0$ the overlap equations (2.226) become identical to those found for attractor networks with binary neurons and finite p (hence our choice to insert an extra factor 2 in defining the synapses), with γ replacing the inverse noise level β in the former.

$$\langle g(\boldsymbol{\xi})\rangle_{\boldsymbol{\xi}} = \lim_{N\to\infty}\frac{1}{N}\sum_i g(\boldsymbol{\xi}_i) = 2^{-p}\sum_{\boldsymbol{\xi}\in\{-1,1\}^p} g(\boldsymbol{\xi}).$$

For the simplest non–trivial choice, $A_{\mu\nu} = \delta_{\mu\nu}$ (i.e., $J_{ij} = (2/N)\sum_\mu \xi_i^\mu \xi_j^\mu$, as in the Hopfield [Hop82] model) equation (2.226) yields the familiar pure and mixture state solutions. For $T = 0$ we find a continuous phase transition from non–recall to pure states of the form $m_\mu = m\delta_{\mu\nu}$ (for some ν) at $\gamma_c = 1$. For $T > 0$ we have in (2.226) an additional Gaussian noise, absent in the models with binary neurons. Again the pure states are the first non–trivial solutions to enter the stage. Substituting $m_\mu = m\delta_{\mu\nu}$ into (2.226) gives

$$m = \int Dz\ \tanh[\gamma(m + z\sqrt{T})]. \tag{2.227}$$

Writing (2.227) as $m^2 = \gamma m\int_0^m dk[1 - \int Dz\ \tanh^2[\gamma(k + z\sqrt{T})]] \le \gamma m^2$, reveals that $m = 0$ as soon as $\gamma < 1$. A continuous transition to an $m > 0$ state occurs when

$$\gamma^{-1} = 1 - \int Dz\ \tanh^2[\gamma z\sqrt{T}].$$

A parametrization of this transition line in the $(\gamma, T)-$plane is given by

$$\gamma^{-1}(x) = 1 - \int Dz\ \tanh^2(zx),$$
$$T(x) = x^2/\gamma^2(x), \qquad x \ge 0.$$

Discontinuous transitions away from $m = 0$ (for which there is no evidence) would have to be calculated numerically. For $\gamma = \infty$ we get the equation $m = \mathrm{erf}[m/\sqrt{2T}]$, giving a continuous transition to $m > 0$ at $T_c = 2/\pi \approx 0.637$. Alternatively the latter number can also be found by taking $\lim_{x\to\infty} T(x)$ in the above parametrization:

$$T_c(\gamma = \infty) = \lim_{x\to\infty} x^2[1 - \int Dz\ \tanh^2(zx)]^2$$
$$= \lim_{x\to\infty}[\int Dz\ \frac{\partial}{\partial z}\tanh(zx)]^2 = [2\int Dz\ \delta(z)]^2 = 2/\pi.$$

Let us now turn to dynamics. It follows from (2.226) that the 'natural' initial conditions for $\overline{u}_{\boldsymbol{\eta}}$ and $\Sigma_{\boldsymbol{\eta}}$ are of the form: $\overline{u}_{\boldsymbol{\eta}}(0) = \boldsymbol{\eta}\cdot\mathbf{k}_0$ and $\Sigma_{\boldsymbol{\eta}}(0) = \Sigma_0$ for all $\boldsymbol{\eta}$. Equivalently:

$$t = 0: \qquad \rho_{\boldsymbol{\eta}}(u) = [2\pi\Sigma_0^2]^{-\frac{1}{2}}\mathrm{e}^{-\frac{1}{2}[u-\boldsymbol{\eta}\cdot\mathbf{k}_0]^2/\Sigma_0^2},$$
$$\mathbf{k}_0 \in \mathbb{R}^p,\ \Sigma_0 \in \mathbb{R}.$$

These would also be the typical and natural statistics if we were to prepare an initial firing state $\{s_i\}$ by hand, via manipulation of the potentials $\{u_i\}$.

For such initial conditions we can simplify the dynamical equation (2.225) to [Coo01, SC00, SC01, CKS05]

$$m_\mu(t) = \langle\ \xi_\mu \int Dz\ \tanh[\gamma(\boldsymbol{\xi}\cdot[\mathbf{k}_0 \mathrm{e}^{-t} \tag{2.228}$$

$$+\ \mathbf{A}\int_0^t ds\ \mathrm{e}^{s-t}\mathbf{m}(s)] + z\sqrt{T+(\Sigma_0^2-T)\mathrm{e}^{-2t}})]\rangle_{\boldsymbol{\xi}}. \tag{2.229}$$

For the special case of the Hopfield synapses, i.e., $A_{\mu\nu}=\delta_{\mu\nu}$, it follows from (2.228) that recall of a given pattern ν is triggered upon choosing $k_{0,\mu}=k_0\delta_{\mu\nu}$ (with $k_0>0$), since then equation (2.228) generates $m_\mu(t)=m(t)\delta_{\mu\nu}$ at any time, with the amplitude $m(t)$ following from

$$m(t) = \int Dz\ \tanh[\gamma[k_0\mathrm{e}^{-t} + \int_0^t ds\ \mathrm{e}^{s-t}m(s) + z\sqrt{T+(\Sigma_0^2-T)\mathrm{e}^{-2t}}]], \tag{2.230}$$

which is the dynamical counterpart of equation (2.227) (to which indeed it reduces for $t\to\infty$).

We finally specialize further to the case where our Gaussian initial conditions are not only chosen to trigger recall of a single pattern $\boldsymbol{\xi}^\nu$, but in addition describe uniform membrane potentials within the sub-lattices, i.e., $k_{0,\mu}=k_0\delta_{\mu\nu}$ and $\Sigma_0=0$, so $\rho_{\boldsymbol{\eta}}(u)=\delta[u-k_0\eta_\nu]$. Here we can derive from (2.230) at $t=0$ the identity $m_0=\tanh[\gamma k_0]$, which enables us to express k_0 as $k_0=(2\gamma)^{-1}\log[(1+m_0)/(1-m_0)]$, and find (2.230) reducing to

$$m(t) = \int Dz\ \tanh[\mathrm{e}^{-t}\log[\frac{1+m_0}{1-m_0}]^{\frac{1}{2}} + \gamma[\int_0^t ds\ \mathrm{e}^{s-t}m(s) + z\sqrt{T(1-\mathrm{e}^{-2t})}]].$$

Compared to the overlap evolution in large networks of binary networks (away from saturation) one can see richer behavior, e.g., non–monotonicity [Coo01, SC00, SC01, CKS05].

2.3.8 Correlation– and Response–Functions

We now turn to correlation functions $C_{ij}(t,t')$ and response functions $G_{ij}(t,t')$. These will become the language in which the partition function methods are formulated, which will enable us to solve the dynamics of recurrent networks in the (complex) regime near saturation (we take $t>t'$):

$$C_{ij}(t,t') = \langle\sigma_i(t)\sigma_j(t')\rangle, \qquad G_{ij}(t,t') = \partial\langle\sigma_i(t)\rangle/\partial\theta_j(t') \tag{2.231}$$

The $\{\sigma_i\}$ evolve in time according to equations of the form (2.186) (binary neurons, sequential updates), (2.205) (binary neurons, parallel updates) or (2.131) (continuous neurons). The θ_i represent thresholds and/or external stimuli, which are added to the local fields in the cases (2.186,2.205), or added to the deterministic forces in the case of a Fokker–Planck equation (2.131).

We retain $\theta_i(t) = \theta_i$, except for a perturbation $\delta\theta_j(t')$ applied at time t' in defining the response function. Calculating averages such as (2.231) requires determining joint probability distributions involving neuron states at different times.

Fluctuation–Dissipation Theorems

For networks of binary neurons with discrete time dynamics of the form $p_{\ell+1}(\boldsymbol{\sigma}) = \sum_{\boldsymbol{\sigma}'} W[\boldsymbol{\sigma};\boldsymbol{\sigma}']\, p_\ell(\boldsymbol{\sigma}')$, the probability of observing a given 'path' $\boldsymbol{\sigma}(\ell') \to \boldsymbol{\sigma}(\ell'+1) \to \ldots \to \boldsymbol{\sigma}(\ell-1) \to \boldsymbol{\sigma}(\ell)$ of successive configurations between step ℓ' and step ℓ is given by the product of the corresponding transition matrix elements (without summation):

$$\begin{aligned}\mathrm{Prob}[\boldsymbol{\sigma}(\ell'),\ldots,\boldsymbol{\sigma}(\ell)] = W[\boldsymbol{\sigma}(\ell);\boldsymbol{\sigma}(\ell-1)]W[\boldsymbol{\sigma}(\ell-1);\\ \boldsymbol{\sigma}(\ell-2)]\ldots W[\boldsymbol{\sigma}(\ell'+1);\boldsymbol{\sigma}(\ell')]p_{\ell'}(\boldsymbol{\sigma}(\ell')).\end{aligned}$$

This allows us to write [Coo01, SC00, SC01, CKS05]

$$\begin{aligned}C_{ij}(\ell,\ell') &= \sum_{\boldsymbol{\sigma}(\ell')} \cdots \sum_{\boldsymbol{\sigma}(\ell)} \mathrm{Prob}[\boldsymbol{\sigma}(\ell'),\ldots,\boldsymbol{\sigma}(\ell)]\sigma_i(\ell)\sigma_j(\ell') \qquad (2.232)\\ &= \sum_{\boldsymbol{\sigma}\boldsymbol{\sigma}'} \sigma_i\sigma'_j W^{\ell-\ell'}[\boldsymbol{\sigma};\boldsymbol{\sigma}']p_{\ell'}(\boldsymbol{\sigma}'),\end{aligned}$$

$$G_{ij}(\ell,\ell') = \sum_{\boldsymbol{\sigma}\boldsymbol{\sigma}'\boldsymbol{\sigma}''} \sigma_i W^{\ell-\ell'-1}[\boldsymbol{\sigma};\boldsymbol{\sigma}'']\left[\frac{\partial}{\partial\theta_j}W[\boldsymbol{\sigma}'';\boldsymbol{\sigma}']\right]p_{\ell'}(\boldsymbol{\sigma}'). \qquad (2.233)$$

From (2.232) and (2.233) it follows that both $C_{ij}(\ell,\ell')$ and $G_{ij}(\ell,\ell')$ will in the stationary state, i.e., upon substituting $p_{\ell'}(\boldsymbol{\sigma}') = p_\infty(\boldsymbol{\sigma}')$, only depend on $\ell - \ell'$: $C_{ij}(\ell,\ell') \to C_{ij}(\ell-\ell')$ and $G_{ij}(\ell,\ell') \to G_{ij}(\ell-\ell')$. For this we do not require detailed balance. Detailed balance, however, leads to a simple relation between the response function $G_{ij}(\tau)$ and the temporal derivative of the correlation function $C_{ij}(\tau)$.

We now turn to equilibrium systems, i.e., networks with symmetric synapses (and with all $J_{ii} = 0$ in the case of sequential dynamics). We calculate the derivative of the transition matrix that occurs in (2.233) by differentiating the equilibrium condition $p_{\mathrm{eq}}(\boldsymbol{\sigma}) = \sum_{\boldsymbol{\sigma}'} W[\boldsymbol{\sigma};\boldsymbol{\sigma}']p_{\mathrm{eq}}(\boldsymbol{\sigma}')$ with respect to external fields:

$$\frac{\partial}{\partial\theta_j}p_{\mathrm{eq}}(\boldsymbol{\sigma}) = \sum_{\boldsymbol{\sigma}'}\{\frac{\partial W[\boldsymbol{\sigma};\boldsymbol{\sigma}']}{\partial\theta_j}p_{\mathrm{eq}}(\boldsymbol{\sigma}') + W[\boldsymbol{\sigma};\boldsymbol{\sigma}']\frac{\partial}{\partial\theta_j}p_{\mathrm{eq}}(\boldsymbol{\sigma}')\}.$$

Detailed balance implies $p_{\mathrm{eq}}(\boldsymbol{\sigma}) = Z^{-1}\mathrm{e}^{-\beta H(\boldsymbol{\sigma})}$ (in the parallel case we simply substitute the appropriate Hamiltonian $H \to \tilde{H}$), giving $\partial p_{\mathrm{eq}}(\boldsymbol{\sigma})/\partial\theta_j = -[Z^{-1}\partial Z/\partial\theta_j + \beta\partial H(\boldsymbol{\sigma})/\partial\theta_j]p_{\mathrm{eq}}(\boldsymbol{\sigma})$, so that

$$\sum_{\boldsymbol{\sigma}'} \frac{\partial W[\boldsymbol{\sigma};\boldsymbol{\sigma}']}{\partial\theta_j} p_{\mathrm{eq}}(\boldsymbol{\sigma}') = \beta\{\sum_{\boldsymbol{\sigma}'} W\left[\boldsymbol{\sigma};\boldsymbol{\sigma}'\right] \frac{\partial H(\boldsymbol{\sigma}')}{\partial\theta_j} p_{\mathrm{eq}}(\boldsymbol{\sigma}') - \frac{\partial H(\boldsymbol{\sigma})}{\partial\theta_j} p_{\mathrm{eq}}(\boldsymbol{\sigma})\},$$

in which the term containing Z drops out. We now get for the response function (2.233) in equilibrium the following result:

$$G_{ij}(\ell) = \tag{2.234}$$
$$\beta \sum_{\boldsymbol{\sigma}\boldsymbol{\sigma}'} \sigma_i W^{\ell-1}\left[\boldsymbol{\sigma};\boldsymbol{\sigma}'\right] \left\{ \sum_{\boldsymbol{\sigma}''} W\left[\boldsymbol{\sigma}';\boldsymbol{\sigma}''\right] \frac{\partial H(\boldsymbol{\sigma}'')}{\partial\theta_j} p_{\mathrm{eq}}(\boldsymbol{\sigma}'') - \frac{\partial H(\boldsymbol{\sigma}')}{\partial\theta_j} p_{\mathrm{eq}}(\boldsymbol{\sigma}') \right\}.$$

The structure of (2.234) is similar to what follows upon calculating the evolution of the equilibrium correlation function (2.232) in a single iteration step:

$$C_{ij}(\ell) - C_{ij}(\ell-1) = \sum_{\boldsymbol{\sigma}\boldsymbol{\sigma}'} \sigma_i W^{\ell-1}\left[\boldsymbol{\sigma};\boldsymbol{\sigma}'\right] \times \tag{2.235}$$
$$\times \{\sum_{\boldsymbol{\sigma}''} W\left[\boldsymbol{\sigma}';\boldsymbol{\sigma}''\right] \sigma_j'' p_{\mathrm{eq}}(\boldsymbol{\sigma}'') - \sigma_j' p_{\mathrm{eq}}(\boldsymbol{\sigma}')\}.$$

Finally we calculate the relevant derivatives of the two Hamiltonians

$$H(\boldsymbol{\sigma}) = -J_{ij}\sigma_i\sigma_j + \theta_i\sigma_i, \qquad \text{and}$$
$$\tilde{H}(\boldsymbol{\sigma}) = -\theta_i\sigma_i - \beta^{-1}\sum_i \log 2\cosh[\beta h_i(\boldsymbol{\sigma})]$$

(with $h_i(\boldsymbol{\sigma}) = J_{ij}\sigma_j + \theta_i$),

$$\frac{\partial H(\boldsymbol{\sigma})}{\partial\theta_j} = -\sigma_j, \qquad \frac{\partial\tilde{H}(\boldsymbol{\sigma})}{\partial\theta_j} = -\sigma_j - \tanh[\beta h_j(\boldsymbol{\sigma})].$$

For sequential dynamics we hereby arrive directly at a *fluctuation–dissipation theorem*. For parallel dynamics we need one more identity (which follows from the definition of the transition matrix in (2.205) and the detailed balance property) to transform the *tanh* occurring in the derivative of $\tilde{H}$:

$$\tanh[\beta h_j(\boldsymbol{\sigma}')] p_{\mathrm{eq}}(\boldsymbol{\sigma}') = \sum_{\boldsymbol{\sigma}''} \sigma_j'' W\left[\boldsymbol{\sigma}'';\boldsymbol{\sigma}'\right] p_{\mathrm{eq}}(\boldsymbol{\sigma}')$$
$$= \sum_{\boldsymbol{\sigma}''} W\left[\boldsymbol{\sigma}';\boldsymbol{\sigma}''\right] \sigma_j'' p_{\mathrm{eq}}(\boldsymbol{\sigma}'').$$

For parallel dynamics ℓ and ℓ' are the real time labels t and t', and we get, with $\tau = t - t'$:

$$G_{ij}(\tau > 0) = -\beta[C_{ij}(\tau+1) - C_{ij}(\tau-1)], \qquad G_{ij}(\tau \le 0) = 0. \tag{2.236}$$

For the continuous-time version (2.186) of sequential dynamics the time t is defined as $t = \ell/N$, and the difference equation (2.235) becomes a differential

equation. For perturbations at time t' in the definition of the response function (2.233) to retain a non-vanishing effect at (re-scaled) time t in the limit $N \to \infty$, they will have to be re-scaled as well: $\delta\theta_j(t') \to N\delta\theta_j(t')$. As a result:

$$G_{ij}(\tau) = -\beta\theta(\tau)\frac{d}{d\tau}C_{ij}(\tau). \tag{2.237}$$

The need to re–scale perturbations in making the transition from discrete to continuous times has the same origin as the need to re-scale the random forces in the derivation of the continuous-time Langevin equation from a discrete-time process. Going from ordinary derivatives to function derivatives (which is what happens in the continuous–time limit), implies replacing Kronecker delta's $\delta_{t,t'}$ by Dirac delta-functions according to $\delta_{t,t'} \to \Delta\delta(t-t')$, where Δ is the average duration of an iteration step. Equations (2.236) and (2.237) are examples of so–called *fluctuation–dissipation theorems* (FDT).

For systems described by a Fokker–Planck equation (2.131) the simplest way to calculate correlation- and response-functions is by first returning to the underlying discrete-time system and leaving the continuous time limit $\Delta \to 0$ until the end. We saw above that for small but finite time-steps Δ the underlying discrete-time process is described by [Coo01, SC00, SC01, CKS05]

$$t = \ell\Delta, \qquad p_{\ell\Delta+\Delta}(\boldsymbol{\sigma}) = [1 + \Delta\mathcal{L}_{\boldsymbol{\sigma}} + \mathcal{O}(\Delta^{\frac{3}{2}})]p_{\ell\Delta}(\boldsymbol{\sigma}),$$

with $\ell = 0, 1, 2, \ldots$ and with the differential operator

$$\mathcal{L}_{\boldsymbol{\sigma}} = -\frac{\partial}{\partial\sigma_i}[f_i(\boldsymbol{\sigma}) - T\frac{\partial}{\partial\sigma_i}].$$

From this it follows that the conditional probability density $p_{\ell\Delta}(\boldsymbol{\sigma}|\boldsymbol{\sigma}', \ell'\Delta)$ for finding state σ at time $\ell\Delta$, given the system was in state $\boldsymbol{\sigma}'$ at time $\ell'\Delta$, must be

$$p_{\ell\Delta}(\boldsymbol{\sigma}|\boldsymbol{\sigma}', \ell'\Delta) = [1 + \Delta\mathcal{L}_{\boldsymbol{\sigma}} + \mathcal{O}(\Delta^{\frac{3}{2}})]^{\ell-\ell'}\delta[\boldsymbol{\sigma} - \boldsymbol{\sigma}']. \tag{2.238}$$

Equation (2.238) will be our main building block. Firstly, we will calculate the correlations:

$$\begin{aligned} C_{ij}(\ell\Delta, \ell'\Delta) &= \langle\sigma_i(\ell\Delta)\sigma_j(\ell'\Delta)\rangle \\ &= \int d\boldsymbol{\sigma}d\boldsymbol{\sigma}'\ \sigma_i\sigma'_j\ p_{\ell\Delta}(\boldsymbol{\sigma}|\boldsymbol{\sigma}', \ell'\Delta)p_{\ell'\Delta}(\boldsymbol{\sigma}') \\ &= \int d\boldsymbol{\sigma}\ \sigma_i[1 + \Delta\mathcal{L}_{\boldsymbol{\sigma}} + \mathcal{O}(\Delta^{\frac{3}{2}})]^{\ell-\ell'} \int\ d\boldsymbol{\sigma}'\ \sigma'_j\delta[\boldsymbol{\sigma} - \boldsymbol{\sigma}']p_{\ell'\Delta}(\boldsymbol{\sigma}') \\ &= \int d\boldsymbol{\sigma}\ \sigma_i[1 + \Delta\mathcal{L}_{\boldsymbol{\sigma}} + \mathcal{O}(\Delta^{\frac{3}{2}})]^{\ell-\ell'}\ [\sigma_j\ p_{\ell'\Delta}(\boldsymbol{\sigma})]\,. \end{aligned}$$

At this stage, we can take the limits $\Delta \to 0$ and $\ell, \ell' \to \infty$, with $t = \ell\Delta$ and $t' = \ell'\Delta$ finite, using $\lim_{\Delta\to 0}[1 + \Delta A]^{k/\Delta} = \mathrm{e}^{kA}$:

$$C_{ij}(t,t') = \int d\boldsymbol{\sigma}\ \sigma_i\ \mathrm{e}^{(t-t')\mathcal{L}_{\boldsymbol{\sigma}}}\left[\sigma_j\ p_{t'}(\boldsymbol{\sigma})\right]. \tag{2.239}$$

Next we turn to the response function. A perturbation applied at time $t' = \ell'\Delta$ to the Langevin forces $f_i(\boldsymbol{\sigma})$ comes in at the transition $\boldsymbol{\sigma}(\ell'\Delta) \to \boldsymbol{\sigma}(\ell'\Delta + \Delta)$. As with sequential dynamics binary networks, the perturbation is re-scaled with the step size Δ to retain significance as $\Delta \to 0$:

$$\begin{aligned}
G_{ij}(\ell\Delta,\ell'\Delta) = \frac{\partial\langle\sigma_i(\ell\Delta)\rangle}{\Delta\partial\theta_j(\ell'\Delta)} &= \frac{\partial}{\Delta\partial\theta_j(\ell'\Delta)}\int d\boldsymbol{\sigma}d\boldsymbol{\sigma}'\ \sigma_i\ p_{\ell\Delta}(\boldsymbol{\sigma}|\boldsymbol{\sigma}',\ell'\Delta)p_{\ell'\Delta}(\boldsymbol{\sigma}') \\
&= \int d\boldsymbol{\sigma}d\boldsymbol{\sigma}'d\boldsymbol{\sigma}''\ \sigma_i\ p_{\ell\Delta}(\boldsymbol{\sigma}|\boldsymbol{\sigma}'',\ell'\Delta+\Delta)\left[\frac{\partial p_{\ell''\Delta+\Delta}(\boldsymbol{\sigma}|\boldsymbol{\sigma}',\ell'\Delta)}{\Delta\partial\theta_j}\right]p_{\ell'\Delta}(\boldsymbol{\sigma}') \\
&= \int d\boldsymbol{\sigma}d\boldsymbol{\sigma}'d\boldsymbol{\sigma}''\ \sigma_i[1+\Delta\mathcal{L}_{\boldsymbol{\sigma}}+\mathcal{O}(\Delta^{\frac{3}{2}})]^{\ell-\ell'-1}\delta[\boldsymbol{\sigma}-\boldsymbol{\sigma}''] \times \\
&\left[\frac{1}{\Delta}\frac{\partial}{\partial\theta_j}[1+\Delta\mathcal{L}_{\boldsymbol{\sigma}''}+\mathcal{O}(\Delta^{\frac{3}{2}})]\delta[\boldsymbol{\sigma}''-\boldsymbol{\sigma}']\right]p_{\ell'\Delta}(\boldsymbol{\sigma}') \\
&= -\int d\boldsymbol{\sigma}d\boldsymbol{\sigma}'d\boldsymbol{\sigma}''\ \sigma_i[1+\Delta\mathcal{L}_{\boldsymbol{\sigma}}+\mathcal{O}(\Delta^{\frac{3}{2}})]^{\ell-\ell'-1} \times \\
&\delta[\boldsymbol{\sigma}-\boldsymbol{\sigma}'']\delta[\boldsymbol{\sigma}''-\boldsymbol{\sigma}'][\frac{\partial}{\partial\sigma'_j}+\mathcal{O}(\Delta^{\frac{1}{2}})]\ p_{\ell'\Delta}(\boldsymbol{\sigma}') \\
&= -\int d\boldsymbol{\sigma}\ \sigma_i[1+\Delta\mathcal{L}_{\boldsymbol{\sigma}}+\mathcal{O}(\Delta^{\frac{3}{2}})]^{\ell-\ell'-1}[\frac{\partial}{\partial\sigma_j}+\mathcal{O}(\Delta^{\frac{1}{2}})]\ p_{\ell'\Delta}(\boldsymbol{\sigma}).
\end{aligned}$$

We take the limits $\Delta \to 0$ and $\ell, \ell' \to \infty$, with $t = \ell\Delta$ and $t' = \ell'\Delta$ finite:

$$G_{ij}(t,t') = -\int d\boldsymbol{\sigma}\ \sigma_i\ \mathrm{e}^{(t-t')\mathcal{L}_{\boldsymbol{\sigma}}}\frac{\partial}{\partial\sigma_j}p_{t'}(\boldsymbol{\sigma}). \tag{2.240}$$

Equations (2.239) and (2.240) apply to arbitrary systems described by Fokker–Planck equations. In the case of conservative forces, i.e., $f_i(\boldsymbol{\sigma}) = -\partial H(\boldsymbol{\sigma})/\partial\sigma_i$, and when the system is in an equilibrium state at time t' so that $C_{ij}(t,t') = C_{ij}(t-t')$ and $G_{ij}(t,t') = G_{ij}(t-t')$, we can take a further step using $p_{t'}(\boldsymbol{\sigma}) = p_{\mathrm{eq}}(\boldsymbol{\sigma}) = Z^{-1}\mathrm{e}^{-\beta H(\boldsymbol{\sigma})}$. In that case, taking the time derivative of expression (2.239) gives

$$\frac{\partial}{\partial\tau}C_{ij}(\tau) = \int d\boldsymbol{\sigma}\ \sigma_i\ \mathrm{e}^{\tau\mathcal{L}_{\boldsymbol{\sigma}}}\mathcal{L}_{\boldsymbol{\sigma}}\left[\sigma_j\ p_{\mathrm{eq}}(\boldsymbol{\sigma})\right].$$

Working out the key term in this expression gives

$$\begin{aligned}
\mathcal{L}_{\boldsymbol{\sigma}}[\sigma_j\ p_{\mathrm{eq}}(\boldsymbol{\sigma})] &= -\sum_i\frac{\partial}{\partial\sigma_i}[f_i(\boldsymbol{\sigma}) - T\frac{\partial}{\partial\sigma_i}][\sigma_j\ p_{\mathrm{eq}}(\boldsymbol{\sigma})] \\
&= T\frac{\partial}{\partial\sigma_j}p_{\mathrm{eq}}(\boldsymbol{\sigma}) - \sum_i\frac{\partial}{\partial\sigma_i}[\sigma_j J_i(\boldsymbol{\sigma})],
\end{aligned}$$

with the components of the probability current density $J_i(\boldsymbol{\sigma}) = [f_i(\boldsymbol{\sigma}) - T\frac{\partial}{\partial\sigma_i}]p_{\rm eq}(\boldsymbol{\sigma})$. In equilibrium, however, the current is zero by definition, so only the first term in the above expression survives. Insertion into our previous equation for $\partial C_{ij}(\tau)/\partial\tau$, and comparison with (2.240) leads to the FDT for continuous systems [Coo01, SC00, SC01, CKS05]:

$$\text{Continuous Dynamics:} \qquad G_{ij}(\tau) = -\beta\theta(\tau)\frac{d}{d\tau}C_{ij}(\tau).$$

We will now calculate the correlation and response functions explicitly, and verify the validity or otherwise of the FDT relations, for attractor networks away from saturation.

Simple Attractor Networks with Binary Neurons

We will consider the continuous time version (2.186) of the sequential dynamics, with the local fields $h_i(\boldsymbol{\sigma}) = J_{ij}\sigma_j + \theta_i$, and the separable interaction matrix (2.202). We already solved the dynamics of this model for the case with zero external fields and away from saturation (i.e., $p \ll \sqrt{N}$). Having non–zero, or even time–dependent, external fields does not affect the calculation much; one adds the external fields to the internal ones and finds the macroscopic laws (2.203) for the overlaps with the stored patterns being replaced by [Coo01, SC00, SC01, CKS05]

$$\dot{\mathbf{m}}(t) = \lim_{N\to\infty}\frac{1}{N}\boldsymbol{\xi}_i\tanh\left[\beta\boldsymbol{\xi}_i\cdot\mathbf{Am}(t)+\theta_i(t)\right] - \mathbf{m}(t), \tag{2.241}$$

Fluctuations in the local fields are of vanishing order in N (since the fluctuations in $\mathbf{m}$ are), so that one can easily derive from the master equation (2.186) the following expressions for spin averages:

$$\frac{d}{dt}\langle\sigma_i(t)\rangle = \tanh\beta[\boldsymbol{\xi}_i\cdot\ \mathbf{Am}(t)+\theta_i(t)] - \langle\sigma_i(t)\rangle, \tag{2.242}$$

$$\begin{aligned} i\neq j:\quad \frac{d}{dt}\langle\sigma_i(t)\sigma_j(t)\rangle &= \tanh\beta[\boldsymbol{\xi}_i\cdot\ \mathbf{Am}(t)+\theta_i(t)]\langle\sigma_j(t)\rangle \\ &+ \tanh\beta[\boldsymbol{\xi}_j\cdot\ \mathbf{Am}(t)+\theta_j(t)]\langle\sigma_i(t)\rangle - 2\langle\sigma_i(t)\sigma_j(t)\rangle. \end{aligned}$$

Correlations at different times are calculated by applying (2.242) to situations where the microscopic state at time t' is known exactly, i.e., where $p_{t'}(\boldsymbol{\sigma}) = \delta_{\boldsymbol{\sigma},\boldsymbol{\sigma}'}$ for some $\boldsymbol{\sigma}'$:

$$\begin{aligned} \langle\sigma_i(t)\rangle|_{\boldsymbol{\sigma}(t')=\boldsymbol{\sigma}'} = \sigma_i'\mathrm{e}^{-(t-t')} + \int_{t'}^{t} ds\ \mathrm{e}^{s-t}\tanh\beta\times \\ \times[\boldsymbol{\xi}_i\cdot\mathbf{Am}(s;\boldsymbol{\sigma}',t')+\theta_i(s)], \end{aligned} \tag{2.243}$$

with $\mathbf{m}(s;\boldsymbol{\sigma}',t')$ denoting the solution of (2.241) following initial condition $\mathbf{m}(t') = \frac{1}{N}\sigma_i'\boldsymbol{\xi}_i$. If we multiply both sides of (2.243) by σ_j' and average over all possible states $\boldsymbol{\sigma}'$ at time t' we get in leading order in N:

$$\langle\sigma_i(t)\sigma_j(t')\rangle = \langle\sigma_i(t')\sigma_j(t')\rangle \mathrm{e}^{-(t-t')} + \int_{t'}^{t} ds\ \mathrm{e}^{s-t}\langle\tanh\beta[\boldsymbol{\xi}_i\cdot\mathbf{Am}(s;\boldsymbol{\sigma}(t'),t')+\theta_i(s)]\sigma_j(t')\rangle.$$

Because of the existence of deterministic laws for the overlaps $\mathbf{m}$ in the $N\to\infty$ limit, we know with probability one that during the stochastic process the actual value $\mathbf{m}(\boldsymbol{\sigma}(t'))$ must be given by the solution of (2.241), evaluated at time t'. As a result we get, with $C_{ij}(t,t') = \langle\sigma_i(t)\sigma_j(t')\rangle$:

$$C_{ij}(t,t') = C_{ij}(t',t')\mathrm{e}^{-(t-t')} + \int_{t'}^{t} ds\ \mathrm{e}^{s-t}\tanh\beta[\boldsymbol{\xi}_i\cdot\mathbf{Am}(s)+\theta_i(s)]\langle\sigma_j(t')\rangle. \tag{2.244}$$

Similarly we get from the solution of (2.242) an equation for the leading order in N of the response functions, by derivation with respect to external fields:

$$\frac{\partial\langle\sigma_i(t)\rangle}{\partial\theta_j(t')} = \beta\theta(t-t')\int_{-\infty}^{t} ds\ \mathrm{e}^{s-t}\left[1-\tanh^2\beta[\boldsymbol{\xi}_i\cdot\mathbf{Am}(s)+\theta_i(s)]\right]\times \times\left[\frac{1}{N}(\boldsymbol{\xi}_i\cdot\mathbf{A}\boldsymbol{\xi}_k)\frac{\partial\langle\sigma_k(s)\rangle}{\partial\theta_j(t')}+\delta_{ij}\delta(s-t')\right], \qquad \text{or}$$

$$G_{ij}(t,t') = \beta\delta_{ij}\theta(t-t')\mathrm{e}^{-(t-t')}\left[1-\tanh^2\beta[\boldsymbol{\xi}_i\cdot\mathbf{Am}(t')+\theta_i(t')]\right] + \beta\theta(t-t')\int_{t'}^{t} ds\ \mathrm{e}^{s-t}\times \times\left[1-\tanh^2\beta[\boldsymbol{\xi}_i\cdot\mathbf{Am}(s)+\theta_i(s)]\right]\frac{1}{N}(\boldsymbol{\xi}_i\cdot\mathbf{A}\boldsymbol{\xi}_k)G_{kj}(s,t'). \tag{2.245}$$

For $t=t'$ we retain in leading order in N only the instantaneous single-site contribution

$$\lim_{t'\uparrow t} G_{ij}(t,t') = \beta\delta_{ij}\left[1-\tanh^2\beta[\boldsymbol{\xi}_i\cdot\mathbf{Am}(t)+\theta_i(t)]\right]. \tag{2.246}$$

This leads to the following ansatz for the scaling with N of the $G_{ij}(t,t')$, which can be shown to be correct by insertion into (2.245), in combination with the correctness at $t=t'$ following from (2.246):

$$i=j: \quad G_{ii}(t,t') = \mathcal{O}(1), \qquad i\neq j: \quad G_{ij}(t,t') = \mathcal{O}(N^{-1})$$

Note that this implies $\frac{1}{N}(\boldsymbol{\xi}_i\cdot\mathbf{A}\boldsymbol{\xi}_k)G_{kj}(s,t') = \mathcal{O}(\frac{1}{N})$. In leading order in N we now find

$$G_{ij}(t,t') = \beta\delta_{ij}\theta(t-t')\mathrm{e}^{-(t-t')}\left[1-\tanh^2\beta[\boldsymbol{\xi}_i\cdot\mathbf{Am}(t')+\theta_i(t')]\right]. \quad (2.247)$$

For those cases where the macroscopic laws (2.241) describe evolution to a stationary state $\mathbf{m}$, obviously requiring stationary external fields $\theta_i(t)=\theta_i$, we can take the limit $t\to\infty$, with $t-t'=\tau$ fixed, in (2.244) and (2.247). Using the $t\to\infty$ limits of (2.242) we subsequently find time translation invariant expressions: $\lim_{t\to\infty}C_{ij}(t,t-\tau)=C_{ij}(\tau)$ and $\lim_{t\to\infty}G_{ij}(t,t-\tau)=G_{ij}(\tau)$, with in leading order in N

$$C_{ij}(\tau) = \tanh\beta[\boldsymbol{\xi}_i\cdot\mathbf{Am}+\theta_i]\tanh\beta[\boldsymbol{\xi}_j\cdot A\mathbf{m}+\theta_j] + \delta_{ij}\mathrm{e}^{-\tau}\left[1-\tanh^2\beta[\boldsymbol{\xi}_i\cdot\mathbf{Am}+\theta_i]\right],$$

$$G_{ij}(\tau) = \beta\delta_{ij}\theta(\tau)\mathrm{e}^{-\tau}\left[1-\tanh^2\beta[\boldsymbol{\xi}_i\cdot\mathbf{Am}+\theta_i]\right],$$

for which the fluctuation–dissipation theorem (2.237) holds [Coo01, SC00, SC01, CKS05]:

$$G_{ij}(\tau) = -\beta\theta(\tau)\frac{d}{d\tau}C_{ij}(\tau).$$

We now turn to the parallel dynamical rules (2.205), with the local fields $h_i(\boldsymbol{\sigma})=J_{ij}\sigma_j+\theta_i$, and the interaction matrix (2.202). As before, having time–dependent external fields amounts simply to adding these fields to the internal ones, and the dynamic laws (2.212) are found to be replaced by

$$\mathbf{m}(t+1) = \lim_{N\to\infty}\frac{1}{N}\boldsymbol{\xi}_i\tanh\left[\beta\boldsymbol{\xi}_i\cdot\mathbf{Am}(t)+\theta_i(t)\right]. \quad (2.248)$$

Fluctuations in the local fields are again of vanishing order in N, and the parallel dynamics versions of equations (2.242), to be derived from (2.205), are found to be

$$\langle\sigma_i(t+1)\rangle = \tanh\beta[\boldsymbol{\xi}_i\cdot A\mathbf{m}(t)+\theta_i(t)], \quad (2.249)$$

$$i\neq j: \qquad \langle\sigma_i(t+1)\sigma_j(t+1)\rangle = \tanh\beta[\boldsymbol{\xi}_i\cdot\mathbf{Am}(t)+\theta_i(t)]\tanh\beta[\boldsymbol{\xi}_j\cdot\mathbf{Am}(t)+\theta_j(t)]. \quad (2.250)$$

With $\mathbf{m}(t;\boldsymbol{\sigma}',t')$ denoting the solution of the map (2.248) following initial condition $\mathbf{m}(t')=\frac{1}{N}\sigma_i'\boldsymbol{\xi}_i$, we immediately get from equations (2.249,2.250) the correlation functions:

$$C_{ij}(t,t) = \delta_{ij}+[1-\delta_{ij}]\tanh\beta[\boldsymbol{\xi}_i\cdot\mathbf{Am}(t-1)+\theta_i(t-1)]\times \tanh\beta[\boldsymbol{\xi}_j\cdot\mathbf{Am}(t-1)+\theta_j(t-1)],$$

$$t>t': \qquad C_{ij}(t,t') = \langle\tanh\beta[\boldsymbol{\xi}_i\cdot\mathbf{Am}(t-1;\boldsymbol{\sigma}(t'),t')+\theta_i(t-1)]\sigma_j(t')\rangle = \tanh\beta[\boldsymbol{\xi}_i\cdot\mathbf{Am}(t-1)+\theta_i(t-1)]\tanh\beta[\boldsymbol{\xi}_j\cdot\mathbf{Am}(t'-1)+\theta_j(t'-1)].$$

From (2.249) also follow equations determining the leading order in N of the response functions $G_{ij}(t,t')$, by derivation with respect to the external fields $\theta_j(t')$:

$$
\begin{aligned}
&t' > t-1 : G_{ij}(t,t') = 0,\\
&t' = t-1 : G_{ij}(t,t') = \beta\delta_{ij}\left[1-\tanh^2\beta[\boldsymbol{\xi}_i\cdot\mathbf{Am}(t-1)+\theta_i(t-1)]\right],\\
&t' < t-1 : G_{ij}(t,t') = \beta\left[1-\tanh^2\beta[\boldsymbol{\xi}_i\cdot\mathbf{Am}(t-1)+\theta_i(t-1)]\right]\times\\
&\times\tfrac{1}{N}(\boldsymbol{\xi}_i\cdot\mathbf{A}\boldsymbol{\xi}_k)G_{kj}(t-1,t').
\end{aligned}
$$

It now follows iteratively that all off–diagonal elements must be of vanishing order in N: $G_{ij}(t,t-1) = \delta_{ij}G_{ii}(t,t-1) \;\rightarrow\; G_{ij}(t,t-2) = \delta_{ij}G_{ii}(t,t-2) \;\rightarrow\; \ldots$, so that in leading order

$$
G_{ij}(t,t') = \beta\delta_{ij}\delta_{t,t'+1}\left[1-\tanh^2\beta[\boldsymbol{\xi}_i\cdot\mathbf{Am}(t')+\theta_i(t')]\right].
$$

For those cases where the macroscopic laws (2.248) describe evolution to a stationary state $\mathbf{m}$, with stationary external fields, we can take the limit $t\to\infty$, with $t-t'=\tau$ fixed above. We find time translation invariant expressions: $\lim_{t\to\infty}C_{ij}(t,t-\tau) = C_{ij}(\tau)$ and $\lim_{t\to\infty}G_{ij}(t,t-\tau) = G_{ij}(\tau)$, with in leading order in N:

$$
\begin{aligned}
C_{ij}(\tau) &= \tanh\beta[\boldsymbol{\xi}_i\cdot\mathbf{Am}+\theta_i]\tanh\beta[\boldsymbol{\xi}_j\cdot\mathbf{Am}+\theta_j]\\
&\quad+\delta_{ij}\delta_{\tau,0}\left[1-\tanh^2\beta[\boldsymbol{\xi}_i\cdot\mathbf{Am}+\theta_i]\right],\\
G_{ij}(\tau) &= \beta\delta_{ij}\delta_{\tau,1}\left[1-\tanh^2\beta[\boldsymbol{\xi}_i\cdot\mathbf{Am}+\theta_i]\right],
\end{aligned}
$$

obeying the Fluctuation-Dissipation Theorem (2.236):

$$
G_{ij}(\tau>0) = -\beta[C_{ij}(\tau+1)-C_{ij}(\tau-1)].
$$

Graded–Response Neurons with Uniform Synapses

Let us finally find out how to calculate correlation and response function for the simple network (2.132) of graded–response neurons, with (possibly time–dependent) external forces $\theta_i(t)$, and with uniform synapses $J_{ij}=J/N$ [Coo01, SC00, SC01, CKS05]:

$$
\dot{u}_i(t) = \frac{J}{N}\sum_j g[\gamma u_j(t)] - u_i(t) + \theta_i(t) + \eta_i(t). \tag{2.251}
$$

For a given realisation of the external forces and the Gaussian noise variables $\{\eta_i(t)\}$ we can formally integrate (2.251) and find

$$
u_i(t) = u_i(0)\mathrm{e}^{-t} + \int_0^t ds\ \mathrm{e}^{s-t}\left[J\int du\ \rho(u;\mathbf{u}(s))\ g[\gamma u] + \theta_i(s) + \eta_i(s)\right], \tag{2.252}
$$

with the distribution of membrane potentials $\rho(u;\mathbf{u}) = N^{-1}\sum_i \delta[u-u_i]$. The correlation function $C_{ij}(t,t') = \langle u_i(t)u_j(t')\rangle$ immediately follows from (2.252). Without loss of generality we can define $t \geq t'$. For absent external forces (which were only needed in order to define the response function), and upon using $\langle \eta_i(s)\rangle = 0$ and $\langle \eta_i(s)\eta_j(s')\rangle = 2T\delta_{ij}\delta(s-s')$, we arrive at

$$C_{ij}(t,t') = T\delta_{ij}(\mathrm{e}^{t'-t} - \mathrm{e}^{-t'-t}) + \langle \left[u_i(0)\mathrm{e}^{-t} + J\int du\ g[\gamma u]\int_0^t ds\ \mathrm{e}^{s-t}\rho(u;\mathbf{u}(s))\right]\times \times \left[u_j(0)\mathrm{e}^{-t'} + J\int du\ g[\gamma u]\int_0^{t'} ds'\ \mathrm{e}^{s'-t'}\rho(u;\mathbf{u}(s'))\right]\rangle.$$

For $N\to\infty$, however, we know the distribution of potentials to evolve deterministically: $\rho(u;\mathbf{u}(s)) \to \rho_s(u)$ where $\rho_s(u)$ is the solution of (2.219). This allows us to simplify the above expression to

$$N\to\infty: \qquad C_{ij}(t,t') = T\delta_{ij}(\mathrm{e}^{t'-t} - \mathrm{e}^{-t'-t}) \tag{2.253}$$
$$+\ \langle \left[u_i(0)\mathrm{e}^{-t} + J\int du\ g[\gamma u]\int_0^t ds\ \mathrm{e}^{s-t}\rho_s(u)\right]\times \times \left[u_j(0)\mathrm{e}^{-t'} + J\int du\ g[\gamma u]\int_0^{t'} ds'\ \mathrm{e}^{s'-t'}\rho_{s'}(u)\right]\rangle.$$

Next we turn to the response function $G_{ij}(t,t') = \delta\langle u_i(t)\rangle/\delta\xi_j(t')$ (its definition involves functional rather than scalar differentiation, since time is continuous). After this differentiation the forces $\{\theta_i(s)\}$ can be put to zero. Functional differentiation of (2.252), followed by averaging, then leads us to

$$G_{ij}(t,t') = \theta(t-t')\ \delta_{ij}\ \mathrm{e}^{t'-t} - J\int du\ g[\gamma u]\frac{\partial}{\partial u}\times \int_0^t ds\ \mathrm{e}^{s-t}\ \frac{1}{N}\sum_k \lim_{\boldsymbol{\theta}\to\mathbf{0}}\langle\ \delta[u-u_k(s)]\frac{\delta u_k(s)}{\delta\theta_j(t')}\ \rangle.$$

In view of (2.252) we make the self-consistent ansatz $\delta u_k(s)/\delta\xi_j(s') = \mathcal{O}(N^{-1})$ for $k\neq j$. This produces

$$N\to\infty: \qquad G_{ij}(t,t') = \theta(t-t')\ \delta_{ij}\ \mathrm{e}^{t'-t}.$$

Since equation (2.219) evolves towards a stationary state, we can also take the limit $t\to\infty$, with $t-t'=\tau$ fixed, in (2.253). Assuming non–pathological decay of the distribution of potentials allows us to put $\lim_{t\to\infty}\int_0^t ds\,\mathrm{e}^{s-t}\rho_s(u) = \rho(u)$ (the stationary solution of (2.219)), with which we find also (2.253) reducing to time translation invariant expressions for $N\to\infty$, $\lim_{t\to\infty} C_{ij}(t,t-\tau) = C_{ij}(\tau)$ and $\lim_{t\to\infty} G_{ij}(t,t-\tau) = G_{ij}(\tau)$, in which

$$C_{ij}(\tau) = T\delta_{ij}\mathrm{e}^{-\tau} + J^2 \left\{ \int du\ \rho(u) g[\gamma u] \right\}^2 ,$$
$$G_{ij}(\tau) = \theta(\tau)\delta_{ij}\mathrm{e}^{-\tau}.$$

Clearly the leading orders in N of these two functions obey the fluctuation-dissipation theorem:

$$G_{ij}(\tau) = -\beta\theta(\tau)\frac{d}{d\tau}C_{ij}(\tau).$$

As with the binary neuron attractor networks for which we calculated the correlation and response functions earlier, the impact of detailed balance violation (occurring when $A_{\mu\nu} \neq A_{\nu\mu}$ in networks with binary neurons and synapses (2.202), and in all networks with graded–response neurons on the validity of the fluctuation-dissipation theorems, vanishes for $N \to \infty$, provided our networks are relatively simple and evolve to a stationary state in terms of the macroscopic observables (the latter need not necessarily happen. Detailed balance violation, however, would be noticed in the finite size effects [CCV98].